

VISIT...

LANZAROTE
Caliente.COM

МИНИСТЕРСТВО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ УЛЬЯНОВСКИЙ
ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ

П.М. Попов, С.П. Попов

**Верификационные методы анализа
оптимального управления
процессами и системами**

УЛЬЯНОВСК 2001

УДК 658-512.011.56.005:621 ББК
32.965.27 П58

Рецензенты: доктор технических наук, профессор А.А. Романцев, кандидат технических наук, доцент С.Д. Лобанов, кандидат технических наук, доцент В.Г. Тодуров, кандидат технических наук, доцент В.П. Туробов

Научный редактор доктор технических наук, профессор В.В. Варнаков

Одобрено редакционно-издательским советом Ульяновского государственного
технического университета

УДК 658-512.011.56.005:621

Попов П.М., Попов С.П.

Верификационные методы анализа оптимального управления процессами и
системами- Ульяновск: УлГТУ, 2001.- 194 с.

Монография создана по результатам исследования методов анализа на основе сравнительных проверок (верификации) оптимального управления основными системами автоматизации и процессами промышленного производства, в том числе авиационного, а также объектами народного хозяйства, экономическими системами и др. В работе использованы экономико-математические, технологические, эксплуатационные, сравнительные проверки некоторых процессов и систем и их поведение в зависимости от тех или иных воздействий и возмущений. Обобщены и дополнены ряд положений оптимизации управления, с использованием элементов классической высшей математики, теории вероятности, математической статистики, вариационного исчисления, динамического программирования, теории игр и решений, принципа максимума и др.

Монография предназначена для использования в учебном процессе вузов, высших военных учебных заведений и может быть полезна аспирантам, научным работникам, работникам промышленности, занимающимся оптимизацией технических и управленческих решений.

Печатается в авторской редакции.

© П.М.Попов, С.П.Попов, 2001
©Оформление. УлГТУ, 2001

ISBN 5-89146-229-X

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ.....	5
1. ВЕРИФИКАЦИЯ ОПТИМАЛЬНОГО УПРАВЛЕНИЯ ПРОЦЕССАМИ И СИСТЕМАМИ АВТОМАТИЗАЦИИ.....	7
1.1. Задачи синтеза систем автоматизации по оптимальному управлению.....	7
1.2. Использование методов классического вариационного исчисления для анализа оптимального управления процессами и системами	12
1.3. Анализ оптимального управления по принципу максимума	21
1.4. Оптимальное управление в линейных системах по быстродействию по принципу максимума	25
1.5. Анализ оптимального управления в системах методом динамического программирования	32
2. ВЕРИФИКАЦИЯ ОПТИМАЛЬНОГО УПРАВЛЕНИЯ С ПОЗИЦИИ ФУНКЦИОНАЛЬНО- СТОИМОСТНОЙ ИНЖЕНЕРИИ И МАРКОВСКИХ ПРОЦЕССОВ ПРИНЯТИЯ РЕШЕНИЙ	39
2.1. Функциональный подход при оценке оптимального управления процессами и системами	40
2.2. Оценка оптимального управления и его эффективность на основе функционально- стоимостного анализа	42
2.3. Сравнительная характеристика функционального синтеза оптимального управления процессами и системами	50
2.4. Процедуры функционально-стоимостного анализа для оценки оптимального управления процессами и системами	51
2.5. Анализ оптимального управления на основе марковских процессов принятия решений	57
2.6. Схемы принятия оптимальных управленческих решений	62
2.7. Анализ оптимальных решений с использованием теории игр (стохастические игры)	66
3. СРАВНИТЕЛЬНЫЙ АНАЛИЗ ОПТИМАЛЬНОГО УПРАВЛЕНИЯ В ПРОЕКТНО- ТЕХНОЛОГИЧЕСКИХ РЕШЕНИЯХ АВИАЦИОННОГО ПРОИЗВОДСТВА.....	71
3.1. Теория оптимизации в проектных решениях подготовки авиационного производства	73
3.2. Оптимальные структуры математических моделей конструкторско- технологического проектирования	74
3.3. Методы оптимизации проектных решений	76
3.4. Предложения по выбору и реализации методов оптимизации при решении проектных задач	90
3.5. Процедуры выбора критериев оптимальности технологических процессов	92
3.6. Моделирующий алгоритм для анализа оптимальности системы массового Обслуживания.....	99

4.ПРОЦЕДУРЫ ПРИНЯТИЯ ОПТИМАЛЬНОГО РЕШЕНИЯ ПРИ ВЫЧИСЛЕНИИ ЭКСТРЕМУМОВ ДЛЯ АНАЛИЗА СИСТЕМЫ	103
4.1. Экстремум без ограничений как процедура принятия оптимального решения	103
4.2. Анализ оптимального решения системы при вычислении экстремума с ограничениями в виде равенств	107
4.3. Динамическая оптимизация без ограничений	116
4.4. Динамическая оптимизация с ограничениями в форме равенств.....	120
4.5. Динамическая оптимизация в форме неравенств с ограничениями	122
4.6. Анализ задачи оптимального управления без ограничений в форме неравенств	124
4.7. Анализ задачи непрерывного оптимального управления в фиксированные и неопределенные моменты достижения	127
5.ОПТИМИЗАЦИЯ СОСТАВА ЭЛЕКТРОННОЙ СИСТЕМЫ УПРАВЛЕНИЯ ДВИГАТЕЛЯМИ И ПО ИЗМЕНЕНИЯМ ПАРАМЕТРОВ.....	132
5.1. Экономико-математический метод оптимизации решений состава ЭСУД на основе законов классической математики	133
5.2. Оптимизация решений при выборе типа электронной системы управления двигателями	136
5.3. Оптимизация процесса технического обслуживания ЭСУД по схеме «затраты-доход» для организации сервиса	138
5.4. Оптимизация параметров ЭСУД при их изменении	141
5.5. Оптимизация параметров ЭСУД с дополнительными условиями при их изменении	142
6.ОПТИМИЗАЦИЯ ПАРАМЕТРОВ ИЗМЕНЕНИЯ ТЕХНИЧЕСКОГО СОСТОЯНИЯ ЭЛЕКТРОННЫХ СИСТЕМ УПРАВЛЕНИЯ ДВИГАТЕЛЯМИ.....	145
6.1. Определение параметров распределения и линейной оптимизации ЭСУД в составе системы «Автомобиль»	145
6.2. Прогнозирование и оптимизация ресурса до усталостного разрушения ЭСУД	153
6.3. Статистический метод оценки и понятие энтропии технического состояния электронной системы управления двигателями автомобилей	158
6.4. Оптимизация затрат и критерии качества текущего ремонта и обслуживания ЭСУД	164
6.5. Оптимизация пропускной способности зоны текущего ремонта и обслуживания ЭСУД в составе системы «Автомобиль»	176
ЗАКЛЮЧЕНИЕ	188
СПИСОК ЛИТЕРАТУРЫ	189

ВВЕДЕНИЕ

Успешное решение задач, связанных с управлением сложными техническими системами и разработками новых процессов (новых технологических процессов), во многом определяется уровнем развития систем автоматизации, систем автоматизированного проектирования и управления разработками. Сроки внедрения научно-технических достижений в различных отраслях деятельности человека непосредственно связаны с качеством получаемой и анализируемой информации об объекте проектирования (создания), информации достоверной, **оптимально построенной** и обработанной для принятия оптимального управленческого решения по объекту (системе) и др. Качество и оптимальность этой информации приобретает особое значение в авиационной и ракетостроительной отрасли знаний, где каждый эксперимент в ходе разработки изделий (систем) связан с большими временными и экономическими (стоимостными) затратами, а получение полной, но оптимальной и достоверной информации об объекте исследований, позволяет резко уменьшить число испытаний и тем самым сократить сроки внедрения образцов новой техники. Поэтому в последние годы большое внимание уделяется вопросам оптимизации технических решений, характеристик систем на основе **оптимальной информации** об объекте (системе, изделии и т.д.) проектирования и организации. Исследования проектирования и управления на основе различных методов анализа и процедур классической высшей математики и функционально-стоимостной инженерии, с помощью которых может быть достигнута заданная цель при условии минимизации (или максимизации) определенного критерия качества любой системы (объекта, изделия и др.) с позиции **функциональности** и **стоимости** (главных критериев проектирования и управления разработками военной техники), составляет фундаментальную задачу оптимизации и оптимального принятия управленческих решений.

Для решения задачи оптимизации и оптимального управления разработками в первую очередь необходимо определить целевую **функцию** или функцию **стоимостную** оптимизируемых процессов и систем. При этом требуется дать соответствующую форму физическую, сформировать задачу для того, чтобы осуществить перевод физического (или технического) описания процесса или системы на язык математики. Для осуществления эффективного управления разработками необходимо знать его текущее состояние, то есть иметь возможность оценивать состояния на основе процедуры **верификации**. Кроме того, необходимо охарактеризовать процесс или систему с помощью адекватной (аналоговой) модели, зависящей от различных факторов, процедуру которой можно назвать идентификацией процесса или системы. При условии использования **функции стоимости**, функционального состояния и параметров системы (процесса), можно затем определить наилучшее (оптимальное) управление разработками, минимизирующее (или максимизирующее) **функцию стоимости и функциональность**. Таким образом, следует сформулировать

несколько взаимосвязанных задач, решение которых позволит построить (спроектировать) оптимальный процесс или оптимальную систему: Задачу верификационных проверок систем оптимального управления; задачу управления разработками; задачу оценки состояния оптимальных процессов и систем; задачу его стохастического управления; задачу адаптации (или адаптивного управления); задачу оценивания параметров в процессах и системах, и др.

Поэтому в настоящей работе авторы представляют верификационные методы (проверки сравнением) анализа оптимального управления процессами и системами, и предлагают изложение проблемы, начиная с рассмотрения оптимальных систем автоматизации, экономических систем и др., и заканчивают доказательством на сугубо гражданском примере уместного использования этих методов анализа с использованием приемов классической высшей математики.

1. ВЕРИФИКАЦИЯ ОПТИМАЛЬНОГО УПРАВЛЕНИЯ ПРОЦЕССАМИ И СИСТЕМАМИ АВТОМАТИЗАЦИИ

Для анализа оптимального управления процессами и системами автоматизации (в т.ч. - автоматизированного) используем проверки сравнением (верификацию) методами классической высшей математики, которые убедительно показывают и определяют - **оптимально** решение того или иного процесса или системы, и насколько эта процедура анализа сама по себе оптимальна.

Итак, в теории оптимальных систем существует два направления управления:

1. Оптимальное управление в **детерминированных системах**;
2. Оптимальное управление в стохастических системах.

В первом известна структура системы и надо найти оптимальные значения ее числовых параметров, при которых обеспечивается заданный **критерий оптимальности** [19].

Во втором - система считается **полностью неизвестной** и требуется определить ее структуру и параметры так, чтобы она была оптимальной по принятому критерию качества (синтезу системы).

В большинстве практических задач оптимизации управленческих решений возможны значительные отступления от оптимальных параметров и структуры без существенного ухудшения качества системы (объекта) управления. При этом либо упрощают исходные уравнения динамики объекта (системы) управления в случае синтеза оптимальной структуры, **либо** упрощают полученную структуру после решения задачи, по исходным уравнениям объекта. Это позволяет управленцу - проектировщику варьировать структуру и изменять параметры в широких пределах для удовлетворения требований, предъявляемых к системе (объекту) управления, среди которых важное значение имеют требования простоты, качества и надежности.

1.1. Задачи синтеза систем автоматизации по оптимальному управлению

Оптимальными системами управления называют системы автоматического, автоматизированного и другого характера, которые обеспечивают выполнение своей **главной функции** наилучшим (оптимальным) образом, при этом обеспечивая **наивысший эффект** (в том или ином смысле), **качество и надежность** в достижении цели (действия - мысленное представление результата, на достижение которого направлено действие). Эти системы могут быть как с обратными связями, так и без них (рис. 1.1, а, б).

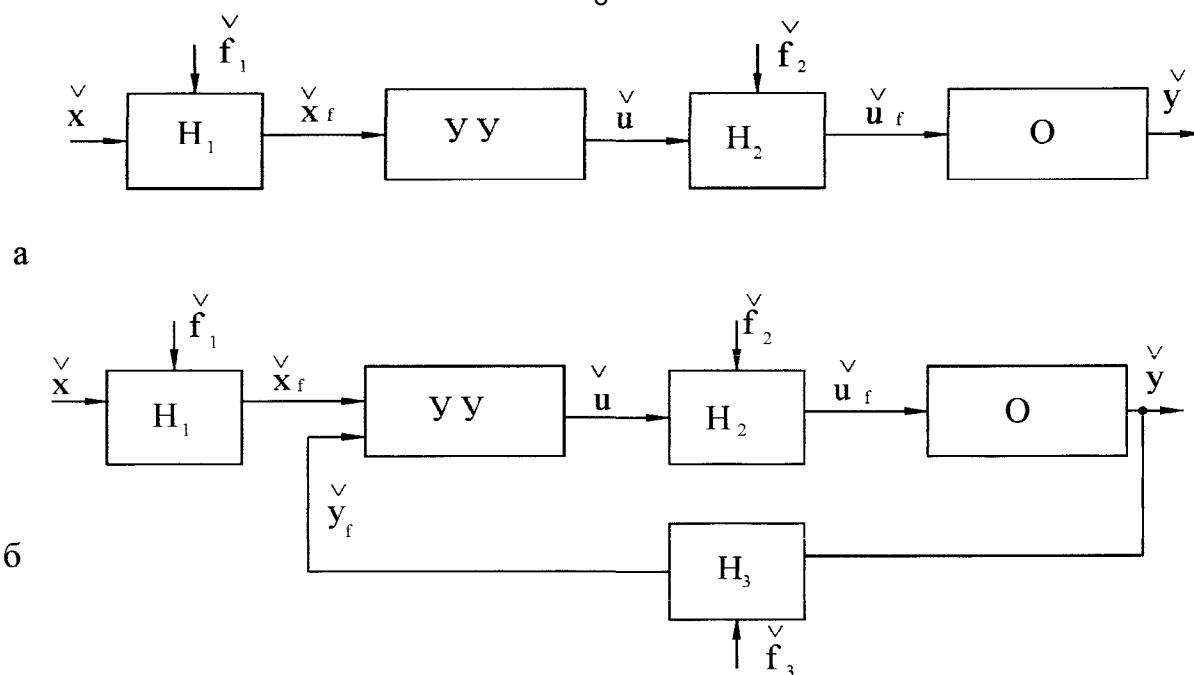


Рис. 1.1. Структурные схемы автоматической системы с оптимальным управлением:

O - объект управления;

UU - управляющее устройство;

$\overset{\vee}{x}$ - задающее воздействие;

$\overset{\vee}{u}$ - управление, формируемое управляющим устройством;

$\overset{\vee}{f}_1, \overset{\vee}{f}_2, \overset{\vee}{f}_3$ - возмущения, действующие на систему

H_1, H_2, H_3 , - операторы, определяющие преобразование функций $\overset{\vee}{x}$, $\overset{\vee}{u}$ и

$\overset{\vee}{y}$ совместно с возмущениями в каналах связи

Задача синтеза оптимальных систем управления включает в себя следующие этапы:

1. Определение математической модели (по возможности электронной модели) объекта управления, то есть определение функциональной зависимости выходной величины y от входного воздействия на объект (или систему) u_f , которая может быть задана различными способами, и в частности системой дифференциальных уравнений.
2. Оценку ограничений как внутренних, присущих физическим процессам в объекте управления, так и наложенных искусственно, извне.
3. Определение желаемого поведения объекта (процесса) управления.

4. Задание определенной цели управления и выбор в соответствии с этой целью критерия оптимальности, характеризующего эффективность управления.

5. Определение стратегии управляющего устройства, иначе говоря, такого алгоритма работы этого устройства, который при указанных выше условиях обеспечивает экстремум критерия оптимальности управления объектом (или процессом), то есть **максимальную эффективность управления**.

6. Схемную реализацию управляющего устройства в соответствии с найденным алгоритмом его работы [19, 22, 3, 9].

Синтез оптимальных систем имеет ряд особенностей по сравнению с синтезом систем другого типа: **во - первых**, он имеет своей целью создание таких систем, у которых используются все их возможности для достижения экстремальных значений наиболее важных показателей качества управления при удовлетворении заданных требований к остальным показателям. **Во-вторых**, при синтезе оптимальных систем энергетические, механические и другие ограничения учитываются как факторы, определяющие возможности систем по реализации экстремальных значений заданных показателей качества управления. Экстремальное значение одного из показателей качества управления оптимальной системы реализуется только в том случае, если управляющее воздействие на объект управления и некоторые его координаты достигают ограничения и остаются определенное время ограниченными в процессе отработки задающего воздействия на систему. При оптимизации одного из качеств системы обычно накладываются ограничения на другие ее свойства. Вопрос об оптимальной системе возникает только тогда, когда потребность в выборе возможно лучшего некоторого показателя качества системы вступает в противоречие с ограниченными ее возможностями. **В-третьих**, качество синтезируемой **оптимальной системы** зависит от правильности выбора **критерия оптимальности**, который характеризует **оптимальность управления**, а достижение экстремума этого критерия является целью управления. Обоснование выбора того или иного критерия оптимальности связано с конкретными технико-экономическими условиями работы системы и в теории оптимальных систем не рассматривается.

Любой критерий оптимальности есть аналитическая оценка оптимизируемого качества системы, зависящая от ее параметров, задающего (x) и возмущающих $(\overset{v}{f}_1, \overset{v}{f}_2, \overset{v}{f}_3)$ воздействий на нее и входного воздействия на объект управления (u) , формируемого управляющим устройством [10]. Следовательно, критерий оптимальности выражается в виде функционала $I(u)$, зависящего от функции управления u , а оптимальное управление u определяется как функция, реализующая экстремум критерия качества, то есть функционала $I(u)$. Например, если к системе (рис. 1.1,б) предъявляется

требование максимальной точности при условии $\overset{\vee}{f}_1 = \overset{\vee}{f}_3 = 0$ то критерием оптимальности может служить интеграл

$$I = \int_0^{\infty} (x - y)^2 dt. \quad (1.1)$$

Так как выходная величина объекта $y(t)$ связана с возмущением $\overset{\vee}{f}_2$ и управлением $\overset{\vee}{u}(t)$ зависимостью, определяемой системой дифференциальных уравнений объекта (или системы, процесса), то формула (1.1) может быть записана в виде

$$I(\overset{\vee}{u}) = \int_0^{\infty} F_0(\overset{\vee}{x}, \overset{\vee}{y}, \overset{\vee}{u}, t) dt. \quad (1.2)$$

Очевидно, что минимальная ошибка будет в системе, управляющее устройство которой формирует такое управление $\overset{\vee}{u}$, при котором выполняется равенство

$$I_{\min}(\overset{\vee}{u}^*) = \min_{\overset{\vee}{u}(t)} \int_0^{\infty} F_0(\overset{\vee}{x}, \overset{\vee}{y}, \overset{\vee}{u}, t) dt = \int_0^{\infty} F_0(\overset{\vee}{x}, \overset{\vee}{y}^*, \overset{\vee}{u}^*, t) dt. \quad (1.3)$$

Такое управление и процессы управления в такой системе управления Γ - $\mathcal{L}$ называются **оптимальными**. Отклонение критерия $J(\overset{\vee}{u})$ от экстремального

значения может служить мерой ухудшения качества системы. Хотя **оптимальная система** решает задачу достижения экстремума одного критерия оптимальности, однако, принятый в качестве этого критерия функционал может представлять любую желаемую комбинацию оценок различных качеств синтезируемой системы, следовательно, в таком случае **оптимизироваться** будет не одно качество, а определенная их совокупность. Как правило, критерии оптимальности строятся так, чтобы цель управления достигалась при достижении минимума критерия.

Задачи синтеза оптимальных систем делятся на два класса задач:

1. Оптимизация программы управления, или определение оптимального

управления $\overset{\vee}{u}(t)$ как функции времени, переводящего объект управления из начального состояния в заданное и реализующего минимум критерия качества, то есть определение алгоритма управляющего устройства системы, схема которой изображена на рис. 1.1, а.

2. Определение закона управления как функции фазовых координат объекта управления $\overset{\vee}{u}(y)$, обеспечивающего движение объекта управления по фазовой траектории, на которой реализуется минимум **критерия**

рия оптимальности, то есть определение алгоритма управляющего устройства замкнутой системы, схема которой изложена на рис.1.1, б. В настоящее время существует несколько направлений классификации **оптимальных систем** управления. Широко используется классификация по **оптимизируемым показателям качества** систем автоматического управления. При этом различают следующие типы систем:

1. Системы управления, оптимальные по быстродействию;
2. Системы управления, оптимальные по расходу ресурсов;
3. Системы управления с минимальной энергией управления;
4. Системы управления с минимальными потерями управления [2,9].

1. Оптимальными по быстродействию называются системы управления, управляющее звено (устройство) которых формирует такое допустимое

управление $u\{t\}$, которое переводит в фазовом пространстве изображающую точку объекта управления из одного заданного состояния $y\{o\}$ в другие $y(t)$ за минимальное время $(T - t_0)$.

2. Системы управления, оптимальные по расходу ресурсов, переводят в фазовом пространстве изображающую объект управления из начального состояния в заданную область S с минимальными затратами ресурсов.

3. Системами управления с минимальной энергией управления называются системы, которые при переводе изображающей точки объекта управления из начального положения в заданное, обеспечивают минимум функцио-

$$I(u) = \int_{t_0}^T u^2(t) dt. \quad (1.4)$$

4. Системы управления с минимальными потерями управления, переводя изображающую точку объекта управления из начального положения в заданное, минимизируют отклонение действительных координат объекта от предписанных значений. К этому типу систем, в частности, относятся **следающие системы** с минимальными ошибками воспроизведения задающего воздействия.

Задача **синтеза оптимальных систем управления**, а точнее, задача определения управляющего звена (устройства) этих систем относится к классу **вариационных задач**. Математической основой методов решения этих задач является **вариационное исчисление**. Для решения вариационных задач наиболее широко используются методы:

- классического вариационного исчисления;
- принцип максимума Понтрягина;
- динамического программирования;
- функционального анализа и др.

1.2. Использование методов классического вариационного исчисления для анализа оптимального управления процессами и системами

Все методы вариационного исчисления позволяют найти условия, при которых достигается экстремум критерия оптимальности, записанного в виде некоторого функционала. Эти условия получаются в виде некоторой системы уравнений относительно уравнения и фазовых координат объекта. Решение этой системы, удовлетворяющее граничным условиям, определяет **оптимальное управление** и оптимальную траекторию изображающей точки объекта управления в его фазовом пространстве [9].

При записи условий существования экстремума функционала используются следующие понятия и определения:

1. Функционал - переменная величина I , зависящая от функций

$$u_i(t), \dot{u}_i(t), y_j(t), \dot{y}_j(t) \quad (i=1,2,\dots,m; j=1,2,\dots,n), \quad (1.5)$$

если каждой из функций (1.5), взятой из некоторого класса этих функций, соответствует определенное значение функционала I . Такая зависимость записывается в виде

$$I = I[u_1(t), u_2(t), \dots, u_m(t); \dot{u}_1(t), \dot{u}_2(t), \dots, \dot{u}_m(t); y_1(t), y_2(t), \dots, y_n(t); \dot{y}_1(t), \dot{y}_2(t), \dots, \dot{y}_n(t)] \quad (1.6)$$

2. Приращение, или вариация, δu_i аргумента $u_i(t)$ функционала (1.6) есть разность функций

$$\delta u_i = u_i(t) - \hat{u}_i(t),$$

где $u_i(t)$ - новая, произвольно выбранная функция из класса функций $u_i(t)$.

3. Близость двух функций характеризуется определенным **порядком близости**. Например, функции $u_i(t)$ и $\hat{u}_i(t)$ близки в смысле близости **нулевого порядка**, если модуль разности $[u_i(t) - \hat{u}_i(t)]$ мал (рис. 1.2,а).

Функции $u_2(t)$ и $u_2(t)$ близки в смысле близости **первого порядка**, если модули

Функционал (1.6) $\left[u_2(t) - \hat{u}_2(t) \right]$ и $u_2^{(1)}(t) - \hat{u}_2^{(1)}(t)$ малы (рис.1.2, б).

такое $\delta > 0$, что

$$\begin{aligned} & \left| I[u_i(t)] - I[u_i^0(t)] \right| < \mathcal{E}; \\ & \left| u_i(t) - u_i^0(t) \right| < \delta; \\ & \left| u_i^{(1)}(t) - u_i^{0(1)}(t) \right| < \delta; \\ & \dots\dots\dots \\ & \left| u_i^{(k)}(t) - u_i^{0(k)}(t) \right| < \delta. \end{aligned}$$

5. Приращение функционала (1.6), соответствующее вариациям аргументов

$$\delta u_1, \delta u_2, \dots, \delta u_m; \delta y_1, \delta y_2, \dots, \delta y_n, \text{ определяется как}$$

$$\begin{aligned} \Delta I = & I(u_1 + \delta u_1, u_2 + \delta u_2, \dots, u_m + \delta u_m; \dot{u}_1 + \delta \dot{u}_1, \dot{u}_2 + \delta \dot{u}_2, \dots, \dot{u}_m; \\ & y_1 + \delta y_1, y_2 + \delta y_2, \dots, y_n + \delta y_n; \dot{y}_1 + \delta \dot{y}_1, \dot{y}_2 + \delta \dot{y}_2, \dots, \dot{y}_n + \delta \dot{y}_n) - \\ & - I(u_1, u_2, \dots, u_m; \dot{u}_1, \dot{u}_2, \dots, \dot{u}_m; y_1, y_2, \dots, y_n; \dot{y}_1, \dot{y}_2, \dots, \dot{y}_n). \end{aligned} \quad (1.7)$$

Если функционал (1.6) имеет в некоторой области непрерывные частные производные второго порядка, то его приращение (1.7) может быть разложено в ряд Тейлора и представлено в виде

$$\begin{aligned}
\Delta I = & \sum_{i=1}^m \left[\frac{\partial I}{\partial u_i} \delta u_i + \frac{\partial I}{\partial \dot{u}_i} \delta \dot{u}_i \right] + \sum_{j=1}^n \left[\frac{\partial I}{\partial y_j} \delta y_j + \frac{\partial I}{\partial \dot{y}_j} \delta \dot{y}_j \right] + \\
& + \frac{1}{2} \left\{ \sum_{i=1}^m \sum_{k=1}^n \left[\frac{\partial^2 I}{\partial u_i \partial u_k} \delta u_i \delta u_k + \frac{\partial^2 I}{\partial u_i \partial \dot{y}_j} \delta u_i \delta \dot{y}_j + \frac{\partial^2 I}{\partial \dot{u}_i \partial \dot{y}_k} \delta \dot{u}_i \delta \dot{y}_k \right] + \right. \\
& + \sum_{i=1}^m \sum_{j=1}^n \left[\frac{\partial^2 I}{\partial u_i \partial y_j} \delta u_i \delta y_j + \frac{\partial^2 I}{\partial u_i \partial \dot{y}_j} \delta u_i \delta \dot{y}_j + \frac{\partial^2 I}{\partial \dot{u}_i \partial \dot{y}_j} \delta \dot{u}_i \delta \dot{y}_j \right] + \\
& \left. + \sum_{j=1}^n \sum_{l=1}^n \left[\frac{\partial^2 I}{\partial y_j \partial y_l} \delta y_j \delta y_l + \frac{\partial^2 I}{\partial y_j \partial \dot{y}_l} \delta y_j \delta \dot{y}_l + \frac{\partial^2 I}{\partial \dot{y}_j \partial \dot{y}_l} \delta \dot{y}_j \delta \dot{y}_l \right] \right\} + o(\rho),
\end{aligned} \tag{1.8}$$

где

$$\rho = \sqrt{(\delta u_1^2 + \delta u_2^2 + \dots + \delta u_m^2 + \delta \dot{u}_1^2 + \delta \dot{u}_2^2 + \dots + \delta \dot{u}_m^2 + \delta y_1^2 + \delta y_2^2 + \dots + \delta y_n^2 + \delta \dot{y}_1^2 + \delta \dot{y}_2^2 + \dots + \delta \dot{y}_n^2)};$$

0 (p) - остаточный член.

Представление приращения функционала (1.6) в форме (1.8) позволяет достаточно просто определить вариации функционала.

6. Если приращение функционала ΔI может быть представлено рядом Тейлора (1.8), то линейная по отношению к вариациям аргументов часть приращения функционала называется **первой вариацией функционала** и записывается в виде

$$\delta I = \sum_{i=1}^m \left[\frac{\partial I}{\partial u_i} \delta u_i + \frac{\partial I}{\partial \dot{u}_i} \delta \dot{u}_i \right] + \sum_{j=1}^n \left[\frac{\partial I}{\partial y_j} \delta y_j + \frac{\partial I}{\partial \dot{y}_j} \delta \dot{y}_j \right].$$

7. Второй **вариацией функционала** (1.6) называется функция

$$\begin{aligned}
\delta^2 I = & \sum_{i=1}^m \sum_{k=1}^n \left[\frac{\partial^2 I}{\partial u_i \partial u_k} \delta u_i \delta u_k + \frac{\partial^2 I}{\partial u_i \partial \dot{y}_j} \delta u_i \delta \dot{y}_j + \frac{\partial^2 I}{\partial \dot{u}_i \partial \dot{y}_k} \delta \dot{u}_i \delta \dot{y}_k \right] + \\
& + \sum_{i=1}^m \sum_{j=1}^n \left[\frac{\partial^2 I}{\partial u_i \partial y_j} \delta u_i \delta y_j + \frac{\partial^2 I}{\partial u_i \partial \dot{y}_j} \delta u_i \delta \dot{y}_j + \frac{\partial^2 I}{\partial \dot{u}_i \partial \dot{y}_j} \delta \dot{u}_i \delta \dot{y}_j \right] + \\
& + \sum_{j=1}^n \sum_{l=1}^n \left[\frac{\partial^2 I}{\partial y_j \partial y_l} \delta y_j \delta y_l + \frac{\partial^2 I}{\partial y_j \partial \dot{y}_l} \delta y_j \delta \dot{y}_l + \frac{\partial^2 I}{\partial \dot{y}_j \partial \dot{y}_l} \delta \dot{y}_j \delta \dot{y}_l \right].
\end{aligned}$$

Необходимым условием существования экстремума непрерывного функционала является равенство нулю его первой вариации

$$\delta I = 0. \tag{1.9}$$

Если при этом достигается минимум функционала, то наряду с выполнением условия (1.9) должно выполняться необходимое условие:

$$\delta^2 I \geq 0,$$

а в случае достижения максимума — условие

$$\delta^2 I \leq 0.$$

Приведенные необходимые условия существования экстремума функционала справедливы, если непрерывный функционал определен на открытом множестве функций (на открытой области некоторого функционального пространства) или если экстремум функционала реализуется функциями, не принадлежащими границе множества, когда функционал определен на замкнутом множестве функций. Особенности определения необходимых условий существования экстремума функционала в случае, когда этот экстремум реализуется функциями, частично и полностью принадлежащими границе множества, на котором этот функционал определен, приведены ниже.

В подавляющем большинстве случаев критерий оптимальности систем автоматического управления записывается в виде интеграла. В частности, он может быть записан так

$$I(\overset{\vee}{u}) = \int_{t_0}^T F_0[\overset{\vee}{u}(t), \overset{\vee}{y}(t)] dt, \quad (1.10)$$

$$F_0[\overset{\vee}{u}(t), \overset{\vee}{y}(t)] = F_0[u_1(t), u_2(t), \dots, u_m(t); \dot{u}_1(t), \dot{u}_2(t), \dots, \dot{u}_m(t); y_1(t),$$

где

$$y_2(t), \dots, y_n(t); \dot{y}_1(t), \dot{y}_2(t), \dots, \dot{y}_n(t)]$$

Если подынтегральная функция

$$F_0[\overset{\vee}{u}(t), \overset{\vee}{y}(t)]$$

непрерывна по совокупно

сти ее аргументов и существуют все ее частные производные до третьего порядка включительно, то необходимые условия экстремума функционала (1.10) записываются в виде системы дифференциальных уравнений Эйлера-Лагранжа

$$\left. \begin{aligned} \frac{\partial F_0}{\partial u_i} - \frac{d}{dt} \cdot \frac{\partial F_0}{\partial \dot{u}_i} &= 0 (i = 1, 2, \dots, m); \\ \frac{\partial F_0}{\partial y_j} - \frac{d}{dt} \cdot \frac{\partial F_0}{\partial \dot{y}_j} &= 0 (j = 1, 2, \dots, n) \end{aligned} \right\} \quad (1.11)$$

Условие (1.11) эквивалентно условию (1.9). Поэтому только на интегральных кривых уравнений Эйлера-Лагранжа, удовлетворяющих граничным условиям

$$\left. \begin{aligned} u_i(t_0), u_i(T) \quad (i = 1, 2, \dots, m); \\ y_j(t_0), y_j(T) \quad (j = 1, 2, \dots, n) \end{aligned} \right\}, \quad (1.12)$$

может реализоваться экстремум (1.10).

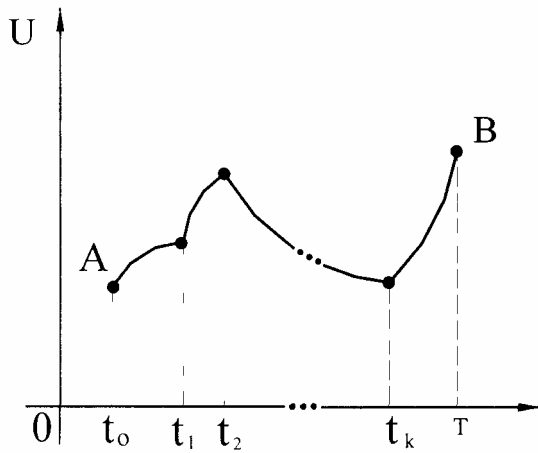


Рис. 1.3. Экстремаль с угловыми точками

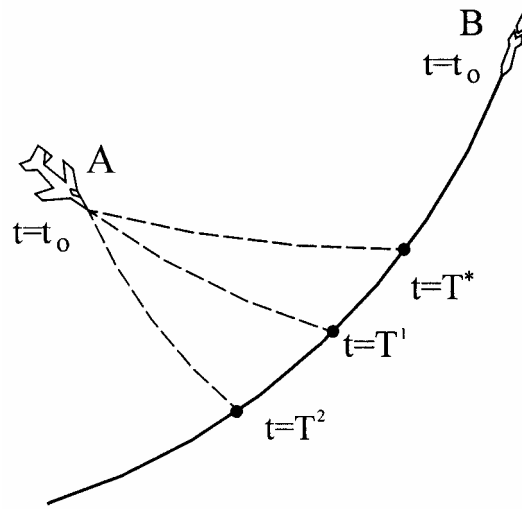


Рис. 1.4. Задача перехвата ракеты В ракетой А за минимальное время

Интегральные кривые уравнения Эйлера-Лагранжа называются **экстремалими**. Экстремали, удовлетворяющие граничным условиям, определяются путем решения краевой задачи. Следует учитывать, что решение не всегда существует, а если и существует, то может быть не единственным. Однако в очень многих задачах синтеза оптимальных систем управления из физического или геометрического смысла задачи достаточно просто устанавливаются существование решения, его единственность и то, что оно реализует минимум критерия оптимальности. В этом случае **экстремали**, удовлетворяющие граничным условиям, есть решение оптимальной задачи.

Если же существует несколько решений уравнений (1.11), удовлетворяющих граничным условиям (1.12), то путем вычисления значений критерия оптимальности на каждом из полученных решений выбирается то из них, на котором критерий достигает минимума.

Экстремум функционала (1.10) может достигаться не на гладких, а на кусочно-гладких экстремалих с конечным числом угловых точек.

Угловыми точками называются точки, в которых экстремали непрерывны

$$u_i^-(t_k) = u_i^+(t_k) ;$$

$$y_j^-(t_k) = y_j^+(t_k) ,$$

а производные от экстремалей терпят разрывы первого рода (рис. 1.3).

$$\dot{u}_i^-(t_k) \neq \dot{u}_i^+(t_k) ;$$

$$\dot{y}_j^-(t_k) \neq \dot{y}_j^+(t_k) ,$$

где t_k - абсцисса k -и угловой точки;

$$u_i^-(t), \dot{u}_i^-, y_j^-(t_k), \dot{y}_j^-(t_k)$$

$$u_i^+(t), \dot{u}_i^+, y_j^+(t_k), \dot{y}_j^+(t_k) \quad \text{соответственно левые и правые пределы экстре}$$

малей и их производных в k -и угловой точке [29].

Если экстремум функционала реализуется на экстремалиях с угловыми точками, которые называются **ломаными экстремалиями**, то в каждой угловой точке должны выполняться условия Вейерштрасса-Эрдмана

$$F_0^- - \sum_{i=1}^m \left(\dot{u}_i \frac{\partial F_0}{\partial \dot{u}_i} \right)^- - \sum_{j=1}^n \left(\dot{y}_j \frac{\partial F_0}{\partial \dot{y}_j} \right)^- = F_0^+ - \sum_{i=1}^m \left(\dot{u}_i \frac{\partial F_0}{\partial \dot{u}_i} \right)^+ - \sum_{j=1}^n \left(\dot{y}_j \frac{\partial F_0}{\partial \dot{y}_j} \right)^+;$$

$$\frac{\partial F_0^-}{\partial \dot{u}_i} = \frac{\partial F_0^+}{\partial \dot{u}_i}; \quad \frac{\partial F_0^-}{\partial \dot{y}_j} = \frac{\partial F_0^+}{\partial \dot{y}_j}.$$

В теории оптимальных систем возникают задачи, когда одна или обе граничные точки экстремалей перемещаются по определенному закону. Например, ракетой А (рис. 1.4) надо управлять так, чтобы уничтожить ракету В за минимальное время. Ракета А запускается с самолета. Очевидно, что в этом случае могут быть заданы только начальные условия (координаты ракет **А** и **В**, значения их скорости, ускорения и т.д. в момент старта) и не могут быть заданы граничные условия, то есть указанные выше параметры в момент встречи двух ракет, так как последние зависят от искомого минимального времени сближения ракет А и В [47].

Подобные задачи в вариационном исчислении называются **задачами с подвижными концами или границами** (а выше рассматривалась **задача с закрепленными концами**). В таких задачах необходимые условия существования экстремума функционала (1.11) должны быть дополнены условиями

$$\left[F_0 - \sum_{j=1}^n \dot{y}_j \frac{\partial F_0}{\partial \dot{y}_j} \right]_{t=t_0} = 0, \quad \left. \frac{\partial F_0}{\partial \dot{y}_j} \right|_{t=t_0} = 0 \quad (j=1,2,\dots,n)$$

$$\left[F_0 - \sum_{j=1}^n \dot{y}_j \frac{\partial F_0}{\partial \dot{y}_j} \right]_{t=T} = 0, \quad \left. \frac{\partial F_0}{\partial \dot{y}_j} \right|_{t=T} = 0 \quad (j=1,2,\dots,n),$$

если не задан закон перемещения конечных точек, или

$$\left. \begin{aligned} & \left[F_0 + \sum_{j=1}^n \left(\dot{\phi}_j - \dot{y}_j \right) \frac{\partial F_0}{\partial \dot{y}_j} \right]_{t=t_0} = 0; \\ & \left[F_0 + \sum_{j=1}^n \left(\dot{\theta}_j - \dot{y}_j \right) \frac{\partial F_0}{\partial \dot{y}_j} \right]_{t=T} = 0 \quad (j=1,2,\dots,n), \end{aligned} \right\} \quad (1.13)$$

где $\phi^{\wedge}$ - закон перемещения концевой точки экстремали $y, \{t_0\}$, ψ - закон перемещения концевой точки экстремали $y, \{T\}$.

Условия (1.13) носят название **условий трансверсальности**. В большинстве случаев при синтезе оптимальных систем возникают задачи минимизации критерия оптимальности при дополнительных условиях, наложенных на координаты объекта управления и на граничные условия. В классическом вариационном исчислении такие задачи получили название **задач на условный экстремум**.

Простейшая задача на условный экстремум формулируется так: требуется исследовать на экстремум функционал (1.10) при условии, что экстремали, на которых реализуется минимум функционала, должны удовлетворять системе дифференциальных уравнений

$$F_j(t, y_1, y_2, \dots, y_n) = 0 \quad (j=1, 2, \dots, n), \quad (1.14)$$

которую называют **системой уравнений связи**.

Эта задача решается путем преобразования функционала (1.10) к виду

$$I^* = \int_{t_0}^T F_0 \left[\overset{\vee}{u}(t), \overset{\vee}{y}(t) \right] dt, \quad (1.15)$$

$$\text{где } F_0^* = F_0 + \sum_{j=1}^n \lambda_j F_j;$$

- неопределенные множители Лагранжа, и исследования нового функционала (1.15) на безусловный экстремум.

Необходимые условия существования **безусловного экстремума** функционала λ_j (1.15) и, следовательно, **условного экстремума** функционала (1.10) имеют вид

$$\left. \begin{aligned} \frac{\partial F_0^*}{\partial u_i} - \frac{d}{dt} \cdot \frac{\partial F_0^*}{\partial \dot{u}_i} &= 0; \\ \frac{\partial F_0^*}{\partial y_i} - \frac{d}{dt} \cdot \frac{\partial F_0^*}{\partial \dot{y}_i} &= 0; \end{aligned} \right\} \quad (1.16)$$

$$F_j(t, y_1, y_2, \dots, y_n) = 0 \quad (i=1, 2, \dots, m; \quad j=1, 2, \dots, n), \quad (1.17)$$

где система (1.16) - система уравнений Эйлера-Лагранжа, а система (1.17) - система дифференциальных уравнений связи.

Условия (1.16) и (1.17) состоят из $(2n+m)$ уравнений относительно

и, следовательно, по

$(2n+m)$ **неизвестных** $u_1, u_2, \dots, u_n; \quad y_1, y_2, \dots, y_n; \quad \dot{u}_1, \dot{u}_2, \dots, \dot{u}_n$

звolyют определить эти неизвестные функции. Решения системы уравнений (1.16) и (1.17) будут содержать $2(2n+m)$ неизвестных постоянных интегрирования. С помощью $2(2n+m)$ граничных условий, заданных для экстремалей $u_i(t)$ и $y_i(t)$ можно определить $2(2n+m)$ произвольных постоянных интегри-

рования. Остальные In постоянных интегрирования находятся путем подбора In незадаанных граничных условий для множителей Лагранжа A_j таким образом, чтобы удовлетворялись граничные условия для функций $u_1(t)$ и $u_l(t)$.

Иногда при решении вариационных задач на **условный экстремум** возникает необходимость выбора из класса кусочно-гладких вектор-функций $u(t)$,

$y(t)$ тех, которые доставляют экстремум функционалу

$$I(\tilde{u}) = q \left[t_0, \check{y}(t_0), T, \check{y}(T) \right] + \int_{t_0}^T F_0 \left[\check{u}(t), \check{y}(t) \right] dt$$

и удовлетворяют уравнениям связи (1.14) и условиям на концах

$$\theta_k \left[t_0, \check{y}(t_0), T, \check{y}(T) \right] = 0 \quad (k = 1, 2, \dots, \rho \leq 2n + 2).$$

Такую задачу называют задачей Больца [27,31].

Применение методов классического вариационного исчисления для решения задач синтеза оптимальных систем связано с рядом трудностей. **Во-первых**, трудности возникают из-за того, что в реальных системах допустимые управляющие воздействия принадлежат замкнутому множеству функций, то есть удовлетворяют условиям

$$|u_1| \leq M_1, |u_2| \leq M_2, \dots, |u_m| \leq M_m,$$

где $M^1, M^2, \dots, M^m$ - заданные константы, и **системы с ограничениями на координаты объекта управления и с ограничениями на возмущения**, и чаще всего наилучшие результаты получаются в том случае, если оптимальные управления выбираются из числа функций, частично или полностью принадлежащих границе этого множества. Например, управляющее напряжение на входе оптимального по быстродействию электропривода, динамика которого описывается дифференциальным уравнением второго порядка, должно изменяться так, как это показано на рис. 1.5. При этом оказывается, что минимум критерия оптимальности, являющегося функционалом от управления $u(t)$, достигается при $|u(t)| = M = \text{const}$, хотя при этих условиях первая вариация функционала не равна нулю (рис. 1.6).

Во-вторых, если оптимальное управление принадлежит к классу кусочно-постоянных функций с конечным числом точек разрывов первого рода, как например, управление, изображенное на рис. 1.5, то это создает значительные вычислительные трудности при определении алгоритма управляющего устройства оптимальной системы [45].

Первая трудность преодолевается путем замены замкнутого множества допустимых **управлений** открытыми. Такая замена может быть осуществлена, в частности, с помощью «функций штрафа» или функций, предложенных Мье-ле.

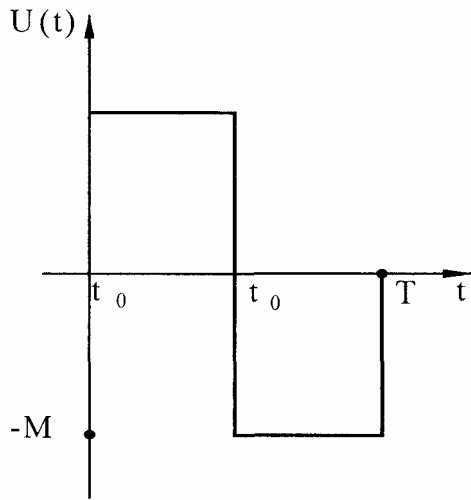


Рис. 1.5. График изменения напряжения на входе оптимального по быстрдействию электропривода

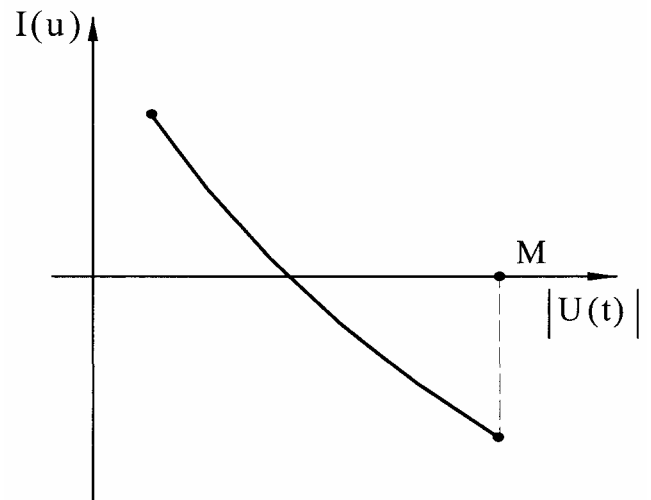


Рис. 1.6. График зависимости величины функционала $I(u)$ от модуля управления $u(t)$

В первом случае при использовании «функций штрафа», в критерий оптимальности вводится дополнительная функция от управления, которая вызывает резкое увеличение критерия, если управление превышает допустимое значение, то есть «штрафует» за «нарушение». Если на управление наложено ограничение (1.19), то «функция штрафа» может быть выбрана в виде (1.20), то есть

$$|u(t)| \leq M = \text{const}, \quad (1.19)$$

то

$$L(u) = \begin{cases} 0 & \text{при } u + M > 0; \\ \alpha(u + M)^2 & \text{при } u + M < 0; \\ 0 & \text{при } M - u > 0; \\ \alpha(M - u)^2 & \text{при } M - u < 0. \end{cases} \quad (1.20)$$

Используя функцию (1.20), можно методами классического вариационного исчисления определять условия существования минимума критерия оптимальности и в случае, когда **оптимальное управление** выбирается из замкнутого множества допустимых управлений.

Во втором случае при использовании функций Мьеле ограничение вида (1.19) учитывается с помощью замены управления $u(t)$ функцией

$$\Phi(\zeta) = \begin{cases} +M & \text{при } \zeta \geq \frac{\Pi}{2}; \\ M \sin \zeta & \text{при } |\zeta| < \frac{\Pi}{2}; \\ -M & \text{при } \zeta \leq -\frac{\Pi}{2}. \end{cases} \quad (1.21)$$

Сведения о методах классического вариационного исчисления, приведенные выше, позволяют дать математическую постановку задачи определения алгоритма управляющего устройства различных типов **оптимальных систем управления**.

1.3. Анализ оптимального управления по принципу максимума

При использовании принципа **максимума** движение объекта управления (системы управления) обычно представляется в виде системы обыкновенных дифференциальных уравнений

$$\dot{y}_j = F_j(y_1, y_2, \dots, y_n; u_1, u_2, \dots, u_m; t), \quad (1.22)$$

где $y_1, y_2, \dots, y_n$ – фазовые координаты;

$u_1, u_2, \dots, u_m$ – управляющие воздействия;

t – время.

Допустимыми считаются управления $u_1, u_2, \dots, u_m$, которые являются непрерывными для всех рассматриваемых t , за исключением конечного числа моментов, где они могут претерпевать разрывы первого рода. На участках непрерывности и в точках разрыва управления они могут принимать лишь конечные значения. Кроме того, на каждое из управлений могут накладываться дополнительные ограничения вида

$$u_{i \min} \leq u_i \leq u_{i \max} \quad (i = 1, 2, \dots, m) \text{ или } R_k(u_1, u_2, \dots, u_m) \leq 0 \quad (k = 1, 2, \dots, 5).$$

Задача **оптимального управления** сводится к отысканию таких управлений, удовлетворяющих наложенным ограничениям, которые одновременно с переводом объекта (системы) из одного положения в другое обеспечивают экстремум выбранного функционала качества. При этом задача может решаться применительно к **автономным и неавтономным системам управления**. Рассмотрим задачу оптимального управления автономной системой, для чего сначала сформулируем определение этой системы. Итак, система называется **автономной**, если правые части дифференциальных уравнений, описывающих ее движение, явно не зависят от времени.

Функционал качества в этом случае выбирается в виде интегрального выражения

$$I = \int_{t_0}^T F_0(y_1, y_2, \dots, y_n; u_1, u_2, \dots, u_m) dt, \quad (1.23)$$

где $F_0(y_1, y_2, \dots, y_n; u_1, u_2, \dots, u_m)$

- заданная функция.

Задача **оптимального управления** по существу сводится к минимизации дополнительной координаты

$$y_0 = \int_{t_0}^T F_0(y_1, y_2, \dots, y_n; u_1, u_2, \dots, u_m) dt,$$

удовлетворяющей условию $y_0=0$ при $t=l_0$. В соответствии с выражением (1.23) к системе (1.22) добавляется еще одно уравнение

$$\dot{y}_0 = F_0(y_1, y_2, \dots, y_n; u_1, u_2, \dots, u_m). \quad (1.24)$$

Далее помимо основной системы уравнений (1.22) и (1.24)

$$\dot{y}_j = F_j(y_1, y_2, \dots, y_n; u_1, u_2, \dots, u_m), \quad (j = 0, 1, \dots, n)$$

рассматривается система сопряженных уравнений для вспомогательных переменных:

$$\begin{aligned} \dot{\psi}_0, \dot{\psi}_1, \dots, \dot{\psi}_n \quad \dot{\psi}_j = & - \left[\frac{\partial F_0(\overset{\vee}{y}, \overset{\vee}{u})}{\partial y_j} \psi_0 + \frac{\partial F_1(\overset{\vee}{y}, \overset{\vee}{u})}{\partial y_j} \psi_1 + \dots + \frac{\partial F_n(\overset{\vee}{y}, \overset{\vee}{u})}{\partial y_j} \psi_n \right] = \\ & = - \sum_{i=0}^n \frac{\partial F_i(\overset{\vee}{y}, \overset{\vee}{u})}{\partial y_j} \psi_i, \quad (j = 0, 1, \dots, n), \end{aligned}$$

$$\overset{\vee}{y} = (y_0, y_1, \dots, y_n);$$

где

$$\overset{\vee}{u} = (u_1, u_2, \dots, u_m).$$

Введение функции Гамильтона

$$H(\overset{\vee}{y}, \overset{\vee}{\psi}, \overset{\vee}{u}) = \psi_0 \dot{y}_0 + \psi_1 \dot{y}_1 + \dots + \psi_n \dot{y}_n = \sum_{i=0}^n F_i(\overset{\vee}{y}, \overset{\vee}{u}) \psi_i$$

позволяет объединить основную и сопряженную системы уравнений одной записью

$$\dot{y}_j = \frac{\partial H(\overset{\vee}{y}, \overset{\vee}{\psi}, \overset{\vee}{u})}{\partial \psi_j}; \quad (1.26)$$

$$\dot{\psi}_j = - \frac{\partial H(\overset{\vee}{y}, \overset{\vee}{\psi}, \overset{\vee}{u})}{\partial y_j}, \quad (j = 0, 1, \dots, n), \quad (1.27)$$

$$\text{где } \overset{\vee}{\psi} = (\psi_0, \psi_1, \dots, \psi_n).$$

Для дальнейшего анализа **оптимального управления по принципу максимума**, рассмотрим его основную теорему, которая основывается на системе дифференциальных уравнений (1.22). Пусть $u(t)$ является управлением,

приводящим изображаемую точку из начального положения $y(t_0)$ в конечное

положение $\overset{\vee}{y}(T)$ а $\overset{\vee}{y}(t)$ - соответствующая этому управлению траектория. Ес-

ли $\overset{\vee}{u}(t)$ оптимально, то найдется такая ненулевая вектор-функция $\overset{\vee}{\phi}(t)$, соот-

ветствующая $u(t)$ и $y(t)$, при которой функция времени, находящийся в заданном интервале (t_0, T) , $H\left(\overset{\vee}{y}, \overset{\vee}{\psi}, \overset{\vee}{u}\right)$ в любой момент максимального значения.

$$H_{\max} = \max_{\overset{\vee}{u} \in M} H\left[\overset{\vee}{y}(t), \overset{\vee}{\psi}(t), \overset{\vee}{u}(t)\right] \quad (1.28)$$

Выражение (1.28) используется для определения функции $u(t)$. Управление будет оптимальным, если оно обеспечивает максимум функции

$$H\left[\overset{\vee}{y}(t), \overset{\vee}{\psi}(t), \overset{\vee}{u}(t)\right] \text{ в любой момент времени.}$$

Следует заметить, что при оптимальном управлении функции $H(t)$ и $\psi_0(t)$ являются постоянными $H(t) = 0, \psi_0(t) \leq 0$. Как видно из уравнений (1.28) принцип максимума устанавливает связь между управлением и координатами основной и сопряженной систем.

Выражения (1.25) и (1.28) позволяют дать геометрическое пояснение принципа максимума.

Величина H является скалярным произведением векторов Ψ и y , поэтому направление движения изображающей точки при оптимальном управлении должно быть таким, чтобы векторы Ψ , y являлись ортогональными. Следовательно, вектор обеспечивает направление движения изображающей точки в фазовом пространстве. Далее, анализируя систему дифференциальных уравнений (1.22), рассмотрим управление **неавтономной системой** и в каком случае управление неавтономной системой может быть оптимальным.

Если в правую часть уравнений системы (1.22) явно входит время t , то задача **оптимального управления** состоит в переводе изображающей точки из начального положения $y(t_0)$ в положение, в котором выполнялись бы условия

$$y_j(t) = y_j(T) \text{ при } j = 1, 2, \dots, n,$$

где $y_j(T)$ - заданные функции времени.

Задача оптимального управления в этом случае может быть сведена к задаче с заданной точкой, если рассматривать движение системы в фазовом пространстве ошибки

$$\overset{\vee}{z}(t) = \overset{\vee}{y}(T) - \overset{\vee}{y}(t),$$

$$\overset{\vee}{z}(t) = [z_1(t), z_2(t), \dots, z_n(t)] ;$$

где

$$\overset{\vee}{y}(T) = [y_1(T), y_2(T), \dots, y_n(T)] .$$

Задача оптимального быстрогодействия является частным случаем задачи с закрепленными концами. Задача состоит в том, чтобы среди всех допустимых

управлений определить такие $\overset{\vee}{u}(t)$, которые переводят изображаемую точку из одного положения в другое за минимальное время. За функционал качества

принимается интегральное выражение вида
$$I = \int_{t_0}^T dt = T - t_0,$$

где T_0 и T - время начала и конца управления.

В случае оптимального быстродействия выражение для функции Гамильтона записывается в следующей форме

$$H = \psi_0 + \sum_{j=1}^n \psi_j F_j(\overset{\vee}{y}, \overset{\vee}{u}).$$

Поскольку во время управления $\phi_0 = const.$ то достаточно рассмотреть

функцию
$$H_1(\overset{\vee}{y}, \overset{\vee}{\psi}, \overset{\vee}{u}) = \sum_{j=1}^n \psi_j F_j(\overset{\vee}{y}, \overset{\vee}{u}) = H(\overset{\vee}{y}, \overset{\vee}{\psi}, \overset{\vee}{u}) - \psi_0.$$

Таким образом, при решении задачи оптимального быстродействия максимум функции Гамильтона будет иметь значение, большее или равное нулю

$$\max_{\overset{\vee}{u} \in M} H_1(\overset{\vee}{y}, \overset{\vee}{\psi}, \overset{\vee}{u}) \geq 0.$$

Как и прежде, вектор $\overset{\vee}{\Psi}$ определяет направление вектора скорости. Однако в этом случае векторы $\overset{\vee}{\Psi}$ и $\overset{\vee}{u}$ могут быть неортогональными.

При определении оптимального управления $\overset{\vee}{u}(t)$ решаются совместно основная (1.26) и сопряженная (1.27) системы уравнений. Для задач с закрепленными концами фазовых траекторий наряду с подбором управления $\overset{\vee}{u}(t)$ максимизирующего функцию $H_1(\overset{\vee}{y}, \overset{\vee}{\phi}, \overset{\vee}{u})$ в каждой точке оптимальной траектории, необходимо знать начальное состояние объекта $y_1(t_0), y_2(t_0), \dots, y_n(t_0)$ и начальное значение вспомогательного вектора $\phi_1(t_0), \phi_2(t_0), \dots, \psi_n(t_0)$. Вектор $\overset{\vee}{y}(t_0)$ задается условиями задачи, а вектор $\overset{\vee}{\phi}(t_0)$ заранее неизвестен. Составляющие вектора, $\psi_j(t_0)$, где $j = 1, 2, \dots, n$, необходимо подобрать таким образом, чтобы оптимальная траектория прошла через заданную конечную точку $y_1(T), y_2(T), \dots, y_n(T)$.

Для задач с подвижными концами граничные положения фазовой точки определяются из условий трансверсальности. Условия трансверсальности для левого и правого концов фазовой траектории определяются ортогональностью

векторов $\overset{\vee}{\Psi}(t_0)$ и $\overset{\vee}{\Psi}(T)$ соответственно касательными векторов многообразий, связывающих начальные и конечные значения фазовых координат

$$\varphi(y_1, y_2, \dots, y_n) = 0;$$

$$\Theta(y_1, y_2, \dots, y_n) = 0.$$

Условия трансверсальности дают дополнительные соотношения, необходимые для определения начальной и конечной точек фазовой траектории.

Из принципа максимума вытекают следующие основные положения теории оптимальных быстрых действий:

1. Оптимальные системы управления являются системами релейного типа;
2. Число переключений оптимального управления конечно и не превышает $(n - 1)$

переключений для систем, характеристические уравнения которых являются единственными отрицательными или нулевыми. В случае же комплексных корней характеристического уравнения число корней может быть больше, чем $(n - 1)$. Вопросы построения оптимального уравнения в нелинейных системах и в системах с ограниченными фазовыми координатами подробно излагаются в работе «Теория оптимизации и расчет систем управления с обратной связью». Задача синтеза замкнутых оптимальных систем состоит в определении управления как функции фазовых координат

$$u = u(y_1, y_2, \dots, y_n).$$

1.4. Оптимальное управление в линейных системах по быстрому действию по принципу максимума

Использование принципа максимума в задачах оптимального быстрого действия приводит к краевой задаче для основной (1.26) и сопряженной (1.27) систем дифференциальных уравнений, решение которой представляет большие трудности. При этом приходится оперировать двумя $(2n)$ неизвестными $y_1,$

$y_2, \dots, y_n; \psi_1, \psi_2, \dots, \psi_n$ и двумя $(2n)$ краевыми условиями, которыми являются

начальные значения векторов $\overset{\vee}{y}(t_0)$ и $\overset{\vee}{\psi}(t_0)$. Начальные условия $y_1(t_0),$

$y_2(t_0), \dots, y_n(t_0)$ известны, а значения $\psi_1(t_0), \psi_2(t_0), \dots, \psi_n(t_0)$ не известны и подбираются из условия удовлетворения граничным условиям на конце оптимальной траектории.

Общих правил подбора значений $\psi_1(t_0), \psi_2(t_0), \dots, \psi_n(t_0)$ не существует. Однако достаточно широкое применение в этих целях получил метод итераций. С помощью принципа максимума сравнительно просто оценивает характер оптимального по быстрому действию управление. Для этого в соответствии с уравнением

$$\dot{y}_j = \sum_{i=1}^n b_{ji} y_i + d_j u, \quad |u| \leq 1 \quad (j = 1, 2, \dots, n)$$

составляется функция Гамильтона $H_1(\overset{\vee}{y}, \overset{\vee}{\psi}, \overset{\vee}{u})$.

Затем определяется уравнение, при котором обеспечивается ее максимум

$$u(t) = \text{sign} \sum_{j=1}^n Q_j \psi_j.$$

Далее определяется, сколько раз изменяется знак управления. Поэтому на примере определения **моментов переключения** на основе стыкования управленческих решений рассмотрим следующую задачу, где расчет алгоритмов управления сводится к определению моментов переключения, которые зависят от многих факторов

$$t_i = f_i \left[\beta_1, \beta_2, K, \beta_n; \overset{\vee}{y}(t_0), \overset{\vee}{y}(T); \overset{\vee}{u} \right],$$

где $\beta_1, \beta_2, \dots, \beta_n$ - параметры объекта управления.

Для определения моментов переключения на практике часто используют метод стыкования решений дифференциальных уравнений, применяя теорему об n интервалах.

Расчет моментов переключения в случае, когда объект описывается дифференциальным уравнением с постоянными коэффициентами

$$y^{(n)} + b_{n-1} \cdot y^{(n-1)} + \dots + b_0 y = Ku_{\max}, \quad (1.29)$$

а начальные и конечные условия представлены векторами

$$y(t_0) = [y_0, y_0^{(1)}, \dots, y_0^{(n-1)}];$$

$$\overset{\vee}{y}(t_n) = [y_n, y_n^{(1)}, \dots, y_n^{(n-1)}],$$

производится по следующей схеме:

1. Находится решение уравнения (1.29)

$$y(t) = C_0 + C_1 e^{-P_1 t} + C_2 e^{-P_2 t} + \dots + C_n e^{-P_n t},$$

где $P_1, P_2, \dots, P_n$ - отрицательные вещественные корни характеристического уравнения.

2. Записываются составляющие вектора системы в фазовом пространстве

на конце последнего интервала уравнения при $t = t_n$

[illegible]

где C_1^n - постоянные интегрирования.

3. Определяются постоянные интегрирования из выражения (1.30).

4. Производится стыкование решений на границе последнего и предпоследнего интервалов

$$\left. \begin{aligned}
& (C_1^{n-1} - C_1^n) e^{-P_1 t_{n-1}} + (C_2^{n-1} - C_2^n) e^{-P_2 t_{n-1}} + \dots + (C_n^{n-1} - C_n^n) e^{-P_n t_{n-1}} = \\
& = \pm 2Ku_{\max}; \\
& -P_1(C_1^{n-1} - C_1^n) e^{-P_1 t_{n-1}} - P_2(C_2^{n-1} - C_2^n) e^{-P_2 t_{n-1}} - \dots - \\
& -P_n(C_n^{n-1} - C_n^n) e^{-P_n t_{n-1}} = 0; \\
& \dots\dots\dots \\
& (-1)^{n-1} P_1^{n-1} (C_1^{n-1} - C_1^n) e^{-P_1 t_{n-1}} + (-1)^{n-1} P_2^{n-1} (C_2^{n-1} - C_2^n) e^{-P_2 t_{n-1}} + \dots + \\
& + (-1)^{n-1} P_n^{n-1} (C_n^{n-1} - C_n^n) e^{-P_n t_{n-1}} = 0.
\end{aligned} \right\} \quad (1.31)$$

5. Система (1.31) решается относительно выражения $(C_1^{n-1} - C_1^n)$.

6. Определяются постоянные интегрирования $C_1^{n-1}, C_2^{n-1}, \dots, C_n^{n-1}$ подстановкой C_i^n ($i=1,2,\dots,n$) значения

7. Стыкуются решения на границе последующих интервалов и определяются выражения $(C_i^{n-2} - C_i^{n-1})$ ($i=1,2,\dots,n$).

Далее производится стыкование решений и исключение постоянных интегрирования до первого интервала.

8. Определяются C_i' из начальных условий путем решения следующей системы уравнения

$$\begin{aligned}
& C_1' + C_2' + \dots + C_n' = y_0 \pm u_{\max}; \\
& -P_1 C_1' - P_2 C_2' - \dots - P_n C_n' = y_0'; \\
& (-1)^{n-1} P_1^{n-1} C_n' + (-1)^{n-1} P_2^{n-1} C_2' + \dots + \\
& + (-1)^{n-1} P_n^{n-1} C_n' = y_0^{(n-1)}.
\end{aligned}$$

9. Приравнением значений C_i' ($i=1,2,\dots,n$) находится система уравнений для определения неизвестных $t_1, t_2, \dots, t_n$.

10. Рассчитываются моменты смены знаков управления. Определив моменты переключения, при необходимости можно перейти к замкнутой форме управления, найдя синтезирующую функцию $u = u[y, y^{(1)}, \dots, y^{(n-1)}]$.

В этом случае управление является не функцией времени, а функцией фазовых координат системы.

В таблице 1.1 приводятся функции оптимального управления для объектов, движение которых описывается дифференциальными уравнениями до третьего порядка включительно. Для этого разберемся, как определить оптимальность управления на основе метода **фазового пространства**, так как метод фазового пространства в сочетании с принципом максимума получил достаточно широкое применение при построении **оптимальных систем**. Здесь движение управляемого объекта описывается дифференциальным [21] уравнением с постоянным коэффициентом

$$\dot{y}_i = \sum_{j=1}^n b_{ij} y_j + d_i u \quad (j=1, 2, \dots, m), \quad (1.32)$$

где $|u| \leq 1$; $y_1, y_2, \dots, y_n$ - координаты объекта, представляющие отклонения от установившегося режима.

Если объект находится в заданном положении, то выполняется условие

$$y_1 = y_2 = \dots = y_n = 0.$$

Следует заметить, что в соответствии с принципом максимума в рассматриваемом случае оптимальное по быстродействию управление является релейным

$$u = \text{sign} \sum_{j=1}^n Q_j \psi_j = \pm 1.$$

Задача синтеза оптимального управления методом фазового пространства сводится к нахождению уравнения **гиперповерхности** переключения в n -мерном фазовом пространстве $\mu(y_1, y_2, \dots, y_n) = 0$ и к определению надлежащего направления переключения реле. **Гиперповерхность** переключения является

односвязной и проходит через начало координат, а управление $u(t)$ теряет свой знак на ее поверхности

$$u = -\text{sign} \mu(y_1, y_2, \dots, y_n).$$

Гиперповерхности переключения в зависимости от структуры и параметров системы, а также внешних воздействий могут быть **нестационарными, квазистационарными и стационарными**. В настоящее время разработаны достаточно эффективные приближенные методы их определения. Для вычисления точек, принадлежащих гиперповерхности переключения, широко применяется метод «попятного движения».

Рассмотрим построение оптимального быстродействия на основе метода фазового пространства системой второго порядка

$$\ddot{y} + \dot{y} = u,$$

где $u = \pm 1$

Ставится задача определения управления, переводящего изображаемую точку фазового пространства из начального положения $y(t_0) = y_0, \dot{y}(t_0) = \dot{y}_0$ в конечное $y(T) = y(T) = 0$ за минимальное время. При этом, согласно теореме об n интервалах, должно быть не более двух интервалов управления.

Уравнение (1.32) представляется в виде

$$\left. \begin{aligned} \dot{y}_1 &= y_2; \\ \dot{y}_2 &= u - y_2 \end{aligned} \right\} \quad (1.33)$$

для построения фазовых траекторий исключается время t из системы (1.33)

$$\frac{dy_1}{dy_2} = \frac{y_2}{u - y_2} \quad (1.34)$$

и находится решение (1.34)

**Таблица 1.1 Дифференциальные уравнения функций
оптимального управления**

Уравнение	Граничные условия	Функция оптимального уравнения
$\ddot{y} = Ku$	$\overset{\vee}{y}(t_0) = 0; \overset{\vee}{y}(T) = y_n$	$\left. \begin{aligned} t_2 &= 2t_1; \\ t_2^2 - 2t_1^2 + \frac{2y_n}{Ku_{max}} &= 0; \\ e^{Pt_2} - 2e^{Pt_1} + 1 &= 0 \end{aligned} \right\}$
$b_2\ddot{y} + b_1\dot{y} = Ku$	$\overset{\vee}{y}(t_0) = 0; \overset{\vee}{y}(T) = y_n$	$t_2 - 2t_1 + \frac{y_n}{Ku_{max}} = 0$
$b_2\ddot{y} + b_1\dot{y} + b_0y = Ku$	$\overset{\vee}{y}(t_0) = 0; \overset{\vee}{y}(T) = y_n$	$\left(1 + \frac{y_n}{Ku_{max}}\right)e^{Pt_2} - 2e^{Pt_1} + 1 = 0$ $(i=1,2)$ $e^{Pt_3} - 2e^{Pt_2} + 2e^{Pt_1} - 1 = 0;$ $t_3 - 2t_2 + 2t_1 = 0$
$b_3\ddot{y} + b_2\dot{y} = Ku$	$\overset{\vee}{y}(t_0) = 0; \overset{\vee}{y}(T) = y_n$	$t_3^2 - 2t_2^2 + 2t_1^2 + 2\frac{2y_n}{Ku_{max}} = 0;$ $e^{Pt_3} - 2e^{Pt_2} + 2e^{Pt_1} - 1 = 0$
$b_3\ddot{y} + b_2\dot{y} + b_1\dot{y} = Ku$	$\overset{\vee}{y}(t_0) = 0; \overset{\vee}{y}(T) = y_n$	$t_3 - 2t_2 + 2t_1 - \frac{y_n}{Ku_{max}} = 0$ $(i=1,2)$
$b_3\ddot{y} + b_2\dot{y} + b_1\dot{y} + b_0y = Ku$	$\overset{\vee}{y}(t_0) = 0; \overset{\vee}{y}(T) = y_n$	$\left(1 - \frac{y_n}{Ku_{max}}\right)e^{Pt_3} - 2e^{Pt_2} + 2e^{Pt_1} - 1 = 0$ $(i=1,2,3)$

$$y_1 = y_{10} + y_{20} - y_2 - u \cdot \ln \frac{u - y_2}{u - y_{20}}.$$

Уравнения фазовых траекторий при этом имеют следующий вид
при $u = +1$

$$y_1 = y_{10} + y_{20} - y_2 - \ln \frac{1 - y_2}{1 - y_{20}}; \quad (1.35)$$

при $u = -1$

$$y_1 = y_{10} - y_{20} + y_2 + \ln \frac{1 + y_2}{1 + y_{20}}. \quad (1.36)$$

Для различных начальных условий можно построить семейство фазовых траекторий (рис. 1.7), соответствующих положительному или отрицательному управляющему воздействию

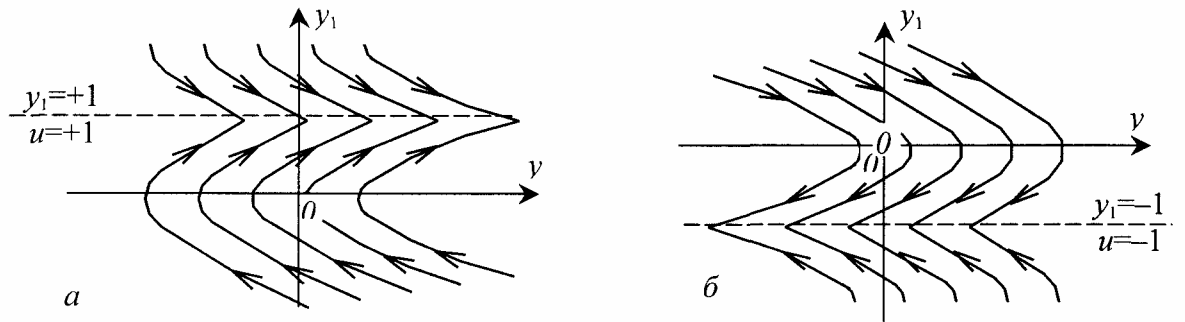


Рис. 1.7. Фазовые траектории системы:

a – при $u = +1$; *б* – при $u = -1$

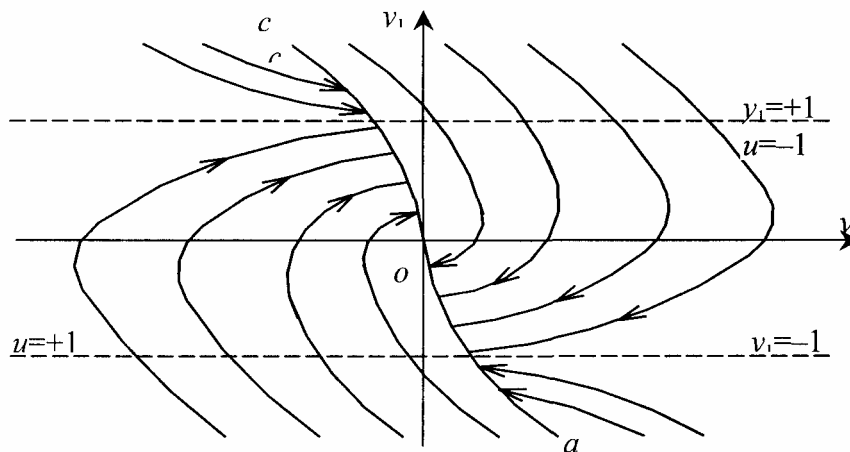


Рис. 1.8. Оптимальные траектории системы

Конечные участки оптимальных фазовых траекторий представляют дуги (кривые **ао** и **со** на рис. 1.8), описываемые уравнениями (1.35) или (1.36) и проходящие через начало координат. На рис. 1.8 видно, что переключение должно происходить при попадании изображающей точки на линию **аос**, которая называется **линией переключения**.

Синтезирующая функция как следует из рис. 1.8, имеет следующий вид

$$u(y_1, y_2) = \begin{cases} +1 & \text{ниже линии } \mathbf{аос} \text{ и на дуге } \mathbf{ао}; \\ -1 & \text{выше линии } \mathbf{аос} \text{ и на дуге } \mathbf{ос}. \end{cases}$$

Уравнение обеих частей линии переключения можно получить соответственно из выражений (1.35) и (1.36), положив

$$\begin{aligned} y_{10} = y_{20} &= 0; \\ y_1 &= -y_2 - \ln(1 - y_2); \\ y_1 &= y_2 + \ln(1 + y_2) \end{aligned}$$

$$\text{или } y_1 = -[|y_2| - \ln(1 + |y_2|)] \text{sign} y_2.$$

Если обозначить через $y_1(t_1)$ значение выходной координаты при нахождении изображающей точки на линии переключения, то сигнал на входе релейного элемента определится выражением

$$z = -[y_1 - y_1(t_1)] = -y_1 - [|y_2| - \ln(1 + |y_2|)] \text{sign} y_2.$$

Полученное уравнение позволяет синтезировать схему оптимального управления (рис. 1.9).

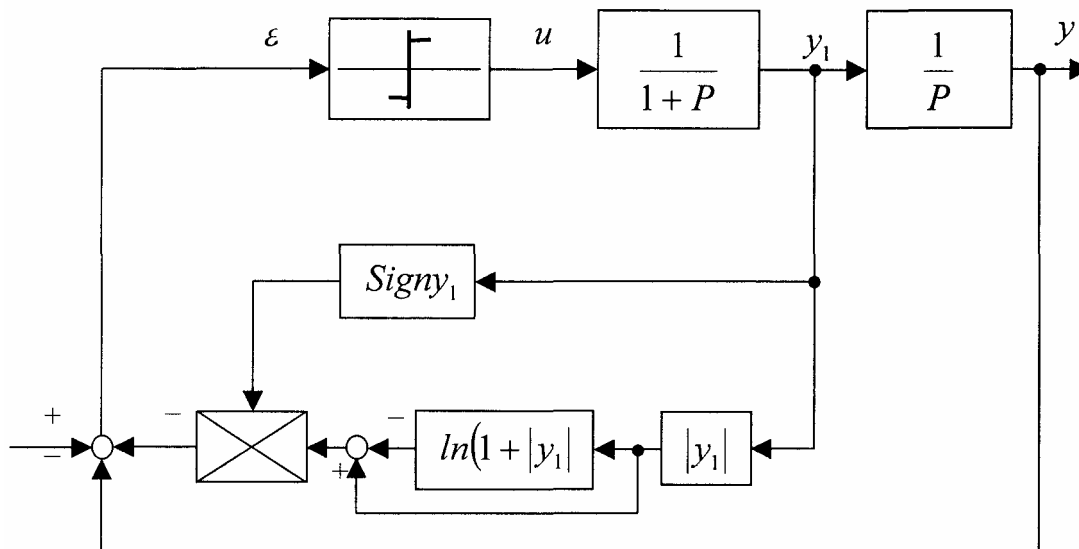


Рис. 1.9. Структурная схема оптимального управления

В ряде случаев возможна реализация оптимального по быстродействию управления с помощью обратных нелинейных и даже линейных обратных связей (1.10).

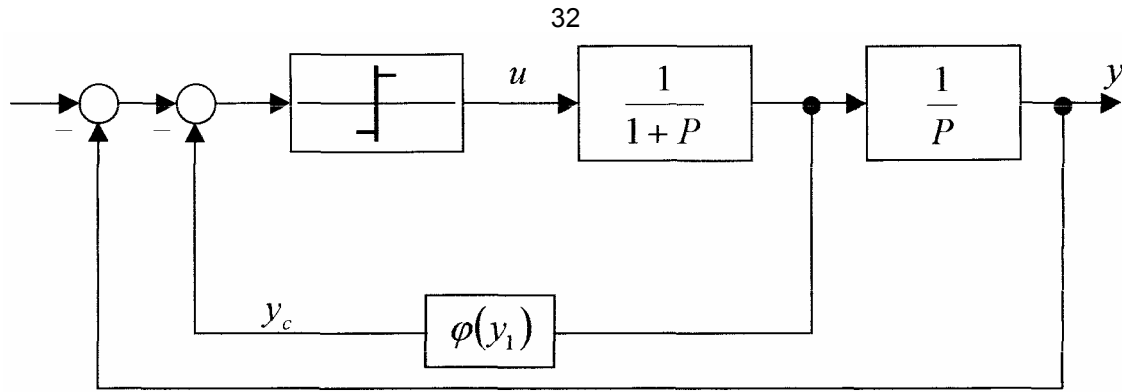


Рис. 1.10. Система оптимального управления с нелинейной обратной связью

Для системы с нелинейной обратной связью, представленной на рис. 1.10, можно записать

$$y_c = [|y_2| - \ln(1 - |y_2|)] \text{sign} y_2.$$

Полученный алгоритм может быть аппроксимирован и реализован с помощью функционального преобразования. При неизменных граничных условиях возможно применение линейных обратных связей. В случае $y_{20} = 0$ коэффициент усиления обратной связи определяется выражением

$$K_c = \frac{y_1}{y_2(t_1)} = 1 - \frac{\ln[1 + |y_2(t_1)|]}{|y_2(t_1)|},$$

где $y_2(t_1) = \sqrt{1 - e^{-|y_{20}|}}$ - скорость изменения выходной координаты в момент смены знаков управления.

В заключение следует отметить, что задача построения оптимальных управлений системами третьего и более высоких порядков методом фазового пространства оказывается сложной. В связи с этим в ряде случаев целесообразно ограничиться синтезом квазиоптимальных систем управления.

1.5. Анализ оптимального управления в системах методом динамического программирования

Суть метода можно пояснить на примере задачи синтеза **оптимального управления объектом** (разработан Р. Беллманом для решения задач оптимального управления) с ограниченными координатами, которое должно переводить его изображающую точку из заданного состояния $y(o)$ в некоторую область (g) фазового пространства за определенное T , минимизируя функционал

$$I(u) = \int_0^T F_0 \left[\overset{\vee}{y}(t), u(t) \right] dt. \quad (1.37)$$

Условия, которым должны удовлетворять фазовые координаты объекта и управляющие воздействия на него, в векторной форме могут быть записаны

$$\left. \begin{aligned} \frac{dy}{dt} &= \overset{\vee}{F} \left(\overset{\vee}{y}, u \right); \\ \overset{\vee}{y}(0) &= \overset{\vee}{y}_0 \in G; \overset{\vee}{y}(T) \in g \subset G; \overset{\vee}{y}(t) \in G; \\ u(t) &\in M; \quad 0 \leq t \leq T, \end{aligned} \right\} \quad (1.38)$$

где G - область фазового пространства, из которой не должна выходить экстремаль $\overset{\vee}{y}(t)$;

M - замкнутое ограниченное множество функций, из которого выбираются кусочно-непрерывные управления $u(t)$;

$\overset{\vee}{y}, \overset{\vee}{F}$ - n -мерные вектора;

u, F_0 - скалярные функции.

Если за начало отсчета взять не $t=0$, а некоторую другую точку t_1 интервала $[0; T]$, а в качестве начальных условий выбрать новую точку $\overset{\vee}{y}(t_1)$ из области G и найти оптимальное управление, минимизирующее функционал

$$\int_{t_1}^T F_0 \left[\overset{\vee}{y}(t), u(t) \right] dt, \quad (1.39)$$

то значение минимума функционала (1.39) будет отличаться от минимума функционала (1.37) при условиях (1.38). Следовательно, минимум функциона-

ла есть функция от начального момента времени t_1 и начальной точки $\overset{\vee}{y}(t_1)$, которую принято обозначать $S \left[t_1, \overset{\vee}{y}(t_1) \right]$. Если функция определена при $t=0$, $\overset{\vee}{y}(0) = \overset{\vee}{y}_0$, то $S \left(0, \overset{\vee}{y}_0 \right)$ есть минимум функционала (1.37).

В основе метода динамического программирования лежит **принцип оптимальности**, сформулированный для широкого круга систем. Обозначив че-

рез $\overset{\vee}{y} \left(0, \overset{\vee}{y}_0 \right)$ оптимальную траекторию в фазовом пространстве, на которой

реализуется минимум функционала (1.37) при условиях (1.38), а через

$\overset{\vee}{y} \left(t_1, 0, \overset{\vee}{y}_0 \right)$ - точку, соответствующую новому началу отсчета t_1 и расположен

ную на оптимальной траектории $\overset{\vee}{y} \left(0, \overset{\vee}{y}_0 \right)$, можно записать такую формулиров-

ку принципа оптимальности: если принять значения $t_1, \check{y}(t_1)$ за начальные, то на интервале $[t_1, T]$ оптимальное уравнение $u[t_1, \check{y}(t_1)]$, на котором реализуется минимум функционала (1.39), совпадает с оптимальным уравнением $u(0, \check{y}_0)$ и, следовательно, участок оптимальной траектории $\check{y}(0, \check{y}_0)$ для задачи с начальными значениями $t=0, \check{y}=0$ на интервале $[t_1, T]$ совпадает с оптимальной траекторией для задачи с начальными значениями $t_1, \check{y}(t_1)$.

Можно дать и другую формулировку принципа оптимальности: оптимальная стратегия не зависит от «предыстории» системы и определяется лишь ее состоянием в рассматриваемый момент времени.

Если известна функция $S(t, \check{y})$, где t - произвольная точка на интервале $[0, T]$, а $\check{y}$ - произвольная точка из области G , то с помощью условий (1.38) не трудно найти оптимальное управление. Однако сложно определить $S(t, \check{y})$ в аналитической форме, поэтому чаще всего эту функцию определяют приближенно. Основой приближенного метода определения $S(t, \check{y})$ служат следующие

положения. Выражение для $S[t_1, \check{y}(t_1)]$ можно записать так

$$S[t_1, \check{y}(t_1)] = \min_{u(t)} \left\{ \int_{t_1}^{t_1 + \Delta t} F_0[\check{y}(t); u(t)] dt + \int_{t_1 + \Delta t}^T F_0[\check{y}(t); u(t)] dt \right\}, \quad (1.40)$$

где t_1 - фиксированный момент времени на интервале $[0, T]$;

Δt - малое положительное число.

Функции $\check{y}(t), u(t)$ связаны условиями

$$\left. \begin{aligned} \frac{d\check{y}}{dt} &= \check{F}(\check{y}, u); \end{aligned} \right\} \quad (1.41a)$$

$$\left. \begin{aligned} \check{y}(t_1) &\in G, \check{y}(T) \in g \subset G, \check{y}(t) \in G; \end{aligned} \right\} \quad (1.41b)$$

$$\left. \begin{aligned} u(t) &\in M, t_1 \leq t \leq T. \end{aligned} \right\} \quad (1.41b)$$

В силу принципа оптимальности поведение $u(t)$ на интервале $[t_1 + \Delta t, T]$

не влияет на величину первого интеграла в выражении (1.40), поэтому $u(t)$ на

этом интервале выбирается так, чтобы минимизировать второй интеграл. Тогда выражение (1.40) можно записать в виде

$$S[t_1, \check{y}(t_1)] = \min_{u(t) \in M} \left\{ \int_{t_1}^{t_1 + \Delta t} F_0[\check{y}(t), u(t)] dt + S[t_1 + \Delta t, \check{y}(t_1 + \Delta t)] \right\}. \quad (1.42)$$

Из формулы (1.42) следует, что управление $u(t)$ на интервале $[t_1, t_1 + \Delta t]$ нужно выбрать так, чтобы минимизировать выражение в фигурных скобках. Поведение $u(t)$ на интервале $[t_1, t_1 + \Delta t]$ влияет не только на величину интеграла в выражении (1.42), но и на величину $S[t_1 + \Delta t, \check{y}(t_1 + \Delta t)]$, так как аргумент этой функции $\check{y}(t_1 + \Delta t)$ в свою очередь является функцией $u(t_1)$ и $\check{y}(t_1)$ в силу уравнения (1.41a).

Трудности нахождения минимума выражения, стоящего в фигурных скобках, заставляют прибегать к допущению, что функции $\check{y}(t)$ и $u(t)$ за время Δt изменяются мало, и их можно считать постоянными. Это допущение позволяет заменить вектор-функцию $\check{F}(\check{y}, u)$ (1.41a) и подынтегральную функцию

$F_0(\check{y}, u)$ их значениями в точке t_1 , а произвольную $\frac{d\check{y}(t)}{dt}$ – конечной размерности $\frac{\check{y}(t_1 + \Delta t) - \check{y}(t_1)}{\Delta t}$. При таких допущениях

$$S[t_1, \check{y}(t_1)] \approx \min_{u(t) \in M} \left\{ F_0[t_1, \check{y}(t_1)] \Delta t + S[t_1 + \Delta t, \check{y}(t_1) + \Delta \check{y}(t_1)] \right\}; \quad (1.43)$$

$$\check{y}(t_1) + \Delta \check{y}(t_1) \approx \check{y}(t_1) + \Delta t \check{F}[t_1, \check{y}(t_1), u(t_1)]. \quad (1.44)$$

Определение $S[t_1, \check{y}(t_1)]$ производится методом попятного движения, то есть начиная с момента времени $t_1 = T - \Delta t$.

Первый шаг. Путем подстановки в выражения (1.43) и (1.44) $t_1 = T - \Delta t$ с учетом того, что $S[T, \check{y}(T)] = 0$ при $\check{y}(T) \in g$, находятся зависимости

$$S[T - \Delta t, \check{y}(T - \Delta t)] = \min_{u(T - \Delta t) \in M} F_0[\check{y}(T - \Delta t), u(T - \Delta t)] \Delta t; \quad (1.45)$$

$$\check{y}(T - \Delta t) + \Delta \check{y}(T - \Delta t) = \check{y}(T - \Delta t) + \Delta t \check{F}[\check{y}(T - \Delta t), u(T - \Delta t)]. \quad (1.46)$$

Минимум правой части выражения (1.45) вычисляется для значений $u(T - \Delta t)$ из M и $\left[\check{y}(T - \Delta t) + \Delta \check{y}(T - \Delta t) \right]$ из g , то есть для значений, удовлетворяющих ограничениям на эти функции. Чаще всего $S\left[T - \Delta t, \check{y}(T - \Delta t)\right]$ получается в виде набора дискретных значений, соответствующих различным значениям $\check{y}(T - \Delta t)$, с которыми связаны определенные значения $u\left[T - \Delta t, \check{y}(T - \Delta t)\right]$.

Второй шаг. Фиксируется момент времени $t = T - 2\Delta t$. С помощью тех же действий, что и на первом шаге, получаются выражения

$$S\left[T - 2\Delta t, \check{y}(T - 2\Delta t)\right] = \min_{u(T - 2\Delta t) \in M} \left\{ F_0\left[\check{y}(T - 2\Delta t), u(T - 2\Delta t)\right] + S\left[T - \Delta t, \check{y}(T - \Delta t)\right] \right\}; \quad (1.47)$$

$$\check{y}(T - 2\Delta t) + \Delta \check{y}(T - 2\Delta t) = \check{y}(T - 2\Delta t) + \Delta \check{F}\left[\check{y}(T - 2\Delta t), u(T - 2\Delta t)\right]. \quad (1.48)$$

Зависимости (1.47) и (1.48) после подстановки в них результатов, полученных на первом шаге, позволяют определить функцию

$$S\left[T - 2\Delta t, \check{y}(T - 2\Delta t)\right] \text{ и управление } u\left[T - 2\Delta t, \check{y}(T - 2\Delta t)\right].$$

В дальнейшем процедура повторяется, и для вычислений могут быть использованы рекуррентные формулы

$$S\left[T - k\Delta t, \check{y}(T - k\Delta t)\right] = \min_{u(T - k\Delta t) \in M} \left\{ F_0\left[\check{y}(T - k\Delta t), u(T - k\Delta t)\right] + S\left[T - (k - 1)\Delta t, \check{y}(T - (k - 1)\Delta t)\right] \right\};$$

$$\check{y}(T - k\Delta t) + \Delta \check{y}(T - k\Delta t) = \check{y}(T - k\Delta t) + \Delta \check{F}\left[\check{y}(T - k\Delta t), u(T - k\Delta t)\right],$$

где k - номер шага.

Вычисления продолжаются до тех пор, пока не будет получена функция

$$S\left[0, \check{y}(0)\right], \text{ где } \check{y}(0) - \text{произвольная точка из множества } G_m. \text{ Если заданная,}$$

начальная точка $\check{y}(0) = \check{y}_0$ из условий (1.38) принадлежит множеству G_m , то,

положив в функции $S\left[0, \overset{\vee}{y}(0)\right] \overset{\vee}{y}(0) = \overset{\vee}{y}_0$, получим $S\left(0, \overset{\vee}{y}\right)$ – минимум функционала (1.37) и $u\left(0, \overset{\vee}{y}\right)$ – оптимальное управление.

Если функция $S\left[t, \overset{\vee}{y}(t)\right]$ имеет непрерывные частные производные по всем своим аргументам, то она может быть найдена из уравнения Беллмана

$$\min_{u(t) \in M} \left\{ \frac{\partial S}{\partial t} + \sum_{i=1}^n \frac{\partial S}{\partial y_i} F_i\left[\overset{\vee}{y}(t), u(t)\right] + F_0\left[\overset{\vee}{y}(t), u(t)\right] \right\} = 0 \quad (1.49)$$

при условиях (1.38) и $S\left[T, \overset{\vee}{y}(T)\right] = 0$.

Существование непрерывно дифференцируемой функции $S\left(t, \overset{\vee}{y}\right)$ – решения уравнения (1.49) – является достаточным **условием оптимальности**. Если существует решение уравнения (1.49), то **соответствующее ему управление** $u\left(t, \overset{\vee}{y}\right)$ будет реализовывать **минимум критерия оптимальности**

(1.37). Требования непрерывной дифференцируемой функции $S\left(t, \overset{\vee}{y}\right)$ серьезно ограничивают использование уравнения (1.49) для синтеза оптимальных систем, так как оно не выполняется во многих даже простых задачах. Однако доказано, что возможно использовать уравнение Беллмана и в случае, когда частные производные от функции $S\left(t, \overset{\vee}{y}\right)$ терпят разрывы на некотором множестве N .

Уравнение (1.49) представляет собой нелинейное дифференциальное уравнение в частных производных, и в настоящее время нет общего метода, позволяющего определить $S\left(t, \overset{\vee}{y}\right)$ и $u\left(t, \overset{\vee}{y}\right)$ в явной аналитической форме.

Каждая новая задача требует особого исследования.

Решение задач синтеза оптимальных систем с помощью динамического программирования численными методами требует такого объема вычислительной работы, которая в настоящее время посильна только современным вычислительным комплексам с большим объемом оперативной памяти (типа RS-6000/42T).

Метод динамического программирования обладает рядом достоинств:

1. В процессе решения задачи **оптимальное управление** определяется как функция фазовых координат объекта управления, что упрощает синтез замкнутых оптимальных систем управления.

2. Метод без принципиальных изменений применим для синтеза как равномерно-оптимальных, так и статистически-оптимальных систем.

3. Принципу оптимальности Беллмана удовлетворяют только те оптимизирующие функции $S\left(t, \overset{\vee}{y}\right)$, которые являются функциями Ляпунова для замкнутой системы. Следовательно, соответствующие этим функциям управления $u\left(t, \overset{\vee}{y}\right)$ и формирующие их управляющие устройства будут обеспечивать как минимум критерия качества, так и устойчивость замкнутых систем.

4. Метод позволяет получать простые для понимания, хорошо осмысливаемые физически алгоритмы решения задач **оптимального управления** на ПЭВМ (ЭВМ).

Таким образом, рассмотренные выше методы анализа оптимального управления и принятия управленческих решений, применительно к процессам и системам автоматического управления, иллюстрируют достоинства и перспективность их использования в сочетании с вычислительными комплексами и ПЭВМ, использование методов синтеза различных типов оптимальных систем и т.д.

2. ВЕРИФИКАЦИЯ ОПТИМАЛЬНОГО УПРАВЛЕНИЯ С ПОЗИЦИИ ФУНКЦИОНАЛЬНО-СТОИМОСТНОЙ ИНЖЕНЕРИИ И МАРКОВСКИХ ПРОЦЕССОВ ПРИНЯТИЯ РЕШЕНИЙ

Сложность производственно-технических и социально-экономических процессов, протекающих в промышленном производстве, постоянно возрастает. Одновременно усиливаются, становятся все более сложными требования, предъявляемые к **управлению** этими процессами.

Важное место в системе современных прогрессивных инструментов управления, содействующих достижению более высокой эффективности производства, бесспорно занимают методы функционально-стоимостного анализа и функционально-стоимостной инженерии.

Функционально-стоимостной анализ - это целенаправленно составленный комплекс методов, сутью которого является поиск и предложение лучшего либо даже принципиально нового решения функций анализируемого объекта с целью повышения эффективности его использования. **Инверсный** функционально-стоимостной анализ (ФСА) является особым случаем применения функционального принципа как инструмента повышения эффективности. **Функционально-стоимостная инженерия** представляет собой применение методологического комплекса стоимостного анализа в сложном и обширном процессе разработки и проектирования нового объекта. Целью этого методологического комплекса является создание **нового объекта**, который эффективно реализовал бы **функции**, выражающие новую общественную потребность, то есть **функциональность** (качество, надежность, эффективность, ремонтпригодность, долговечность).

Конкретно речь идет о системном применении **функционального** принципа при подготовке и создании нового изделия, технологического процесса, системы и другого объекта. Сюда относится также применение функционально-стоимостной инженерии при проектировании крупных материальных или процессных систем, какими являются, например, самолет, ракета, организационные и управленческие **структуры и системы**.

Рассмотрим систему, у которой пространство состояний S содержит конечное число элементов. Пусть S совпадает с множеством целых чисел $S = \{1, 2, \dots, N\}$. Каждому состоянию $i \in S$ соответствует конечное множество K , решений (или альтернатив), элементы которого обозначим $k = 1, 2, \dots, A_i$.

Пространством политик K назовем прямое произведение множеств решений, то есть $K = K_1 * K_2 * \dots * K_n$. Рассматривается **задача** принятия **оптимальных** последовательных управленческих решений, состоящая в выборе решений при наблюдении текущих состояний в моменты $n = 0, 1, 2, \dots$

Если система находится в состоянии $i \in S$ и принимается решение $k \in K_i$, то:

1. Система получает доход r_{ij}^k ;

2. Ее состояние в следующий момент времени определяется вероятностным законом $P_{ij}^k(j \in S)$, где P_{ij}^k - вероятность того, что система управления из состояния i при выборе оптимального решения k попадает в состояние j . Предполагается, что доход r_{ij}^k ограничен при всех $i \in S$ и $k \in K_i$. Кроме того

$$\sum_{j \in S} P_{ij}^k = 1, \quad P_{ij}^k \geq 0 \quad \text{при} \quad i, j \in S, \quad k \in K_i.$$

(2.1)

Рассмотрим процесс управления с переоценкой. Пусть β , $0 \leq \beta \leq 1$ – ко-

эффициент переоценки. Смысл его состоит в том, что единица дохода через время n (например, n дней) будет стоить β^n единиц.

Введение коэффициента переоценки с математической точки зрения ведет к ограниченности суммарного среднего дохода. Зададим начальное распределение

$$a = (a_1, a_2, \dots, a_N),$$

где $\sum_{i \in S} a_i = 1, \quad a_i \geq 0 \quad \text{при} \quad i \in S.$

(2.3)

Тогда система управления описывается неоднородной цепью Маркова с доходами и т.д.

2.1. Функциональный подход при оценке оптимального управления процессами и системами

Для оценки функций оптимального управления необходим подсчет следующих критериальных составляющих:

1. Показателя значения функции (V_i);
2. Величины степени выполнения функции ($^0F_{ij}$);
3. Величины затрат на функцию (C_{ij}).

Показатель значения функции (V_i) определяется в единицах значимости (j_v).

При определении показателя значение функции прежде всего методом попарного сравнения определяется величина коэффициента значения функции (k_i). Далее, с помощью шкалы отклонений определяется показатель значения отдельных функций по следующей формуле

$$\dot{V}_i = \frac{V_i}{\sum_{i=1}^n K_i} \cdot k_i, \quad (2.4)$$

где $\hat{\cdot}$ - общий показатель значения всех функций в совокупности, выражается в баллах (100 или 1000 j_v). [2]

Исходным является $\sum \dot{V}_i = V_i$. (2.5)

Величина степени выполнения функции ($^{\circ}F_y$), или функциональность объекта, определяется с помощью индексирования и математико-статистических методов (дискриминантный анализ, факторный анализ и т.д.), которые исходят из принципа измерения полезности объекта.

Величина затрат на функцию (C_{ij}) определяется аналитически, расчетным путем. Сущность этого метода заключается в экспертном, профессиональном определении доли участия отдельных элементов структуры объекта (системы) в обеспечении отдельных функций. Затраты на функции определяются так, что затраты на отдельные элементы распределяются на соответствующие функции в такой же пропорции, в какой они обеспечивают реализацию данной функции.

Целью функционального анализа, кроме оценки эффективности управления объекта, является определение его **критической функции** как источника резервов для дальнейшего возможного повышения эффективности. Под **критической функцией** понимается та функция, которая по своим параметрам, функциональности ($^{\circ}F_{ij}$) и затратоемкости (C_{ij}) отрицательно отличается от требуемого уровня. Критическими считаются также функции отсутствующие, ошибочные и лишние.

Самым распространенным методом определения критических функций является **метод построения диаграммы оценки функций**. Сущностью этого метода является сравнительный анализ графического изображения результатов. Предметом сравнения отдельных функций являются здесь все три оценки, а именно **показатель значения функции, величина степени выполнения функции, величина затрат на функцию**.

При применении этого метода производится следующая последовательность технологических процедур (процесса):

1. Составляется диаграмма - график оценки функций. Показатель значения функции анализируемого объекта ($V_{1,}$) отражается (в соответствии с размером) в верхней части графика (рис.2.1). Величина степени выполнения функции анализируемого объекта ($^{\circ}F_{ij}$) указывается в соответствующих колонках, отражающих значение функции. Эта графическая форма отражает достижение каждой функцией при ее выполнении общественного оптимума либо его превышения. С графической стороны уровень общественного оптимума тождествен верхней границе колонки, которая одновременно выражает условную оценку значения функции.

Затраты на отдельные функции анализируемого объекта (C) опять в форме колонок чертятся в нижней части графика оценки функций управления против колонок, выражающих значение и степень выполнения соответствующих функций.

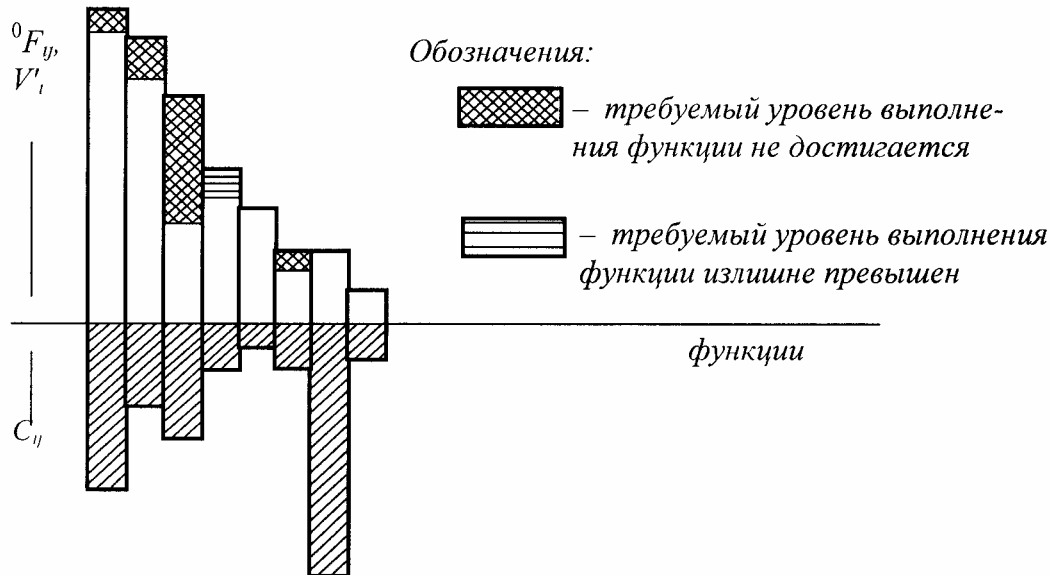


Рис.2.1. Диаграмма оценки функций управления

2. Определяется критическая функция с точки зрения затрат. Путем сравнения показателя значения функций (V_1) с затратами на их обеспечение (C_{ij}) определяется критическая функция. Но это не обязательно самая затратно-емкая функция в комплексе. Это должна быть та функция, затраты на которую не должны быть чрезмерно велики по сравнению с показателем ее значения.

3. Определяется **критическая функция** с потребительской точки зрения. Путем сравнения фактической степени выполнения функций с уровнем общественного оптимума данной функции определяется критическая функция с потребительской точки зрения. Это та функция, которая уровня общественного оптимума либо не достигла, либо его превзошла.

В некоторых методиках и в практике ФСА можно встретиться с так называемыми и **негативными функциями** и т.д.

2.2. Оценка оптимального управления и его эффективности на основе функционально-стоимостного анализа

Одной из важных черт в оптимизации управления и повышения его эффективности на основе функционально-стоимостного анализа является **комплексное рассмотрение эффективности** как органического единства дейст-

венности и целесообразности оцениваемого объекта управления (или процесса управления).

Действенность объекта (процесса) и понимается как его способность трансформировать источники (а—акции) в результаты (r—реакции) по формуле

$$\dot{u} = \frac{r}{a} = \max. \quad (2.6)$$

Если выразить акции и реакции в экономических категориях, то следует говорить об экономической эффективности $\dot{u}_{ek}$, по которой оцениваем трансформацию количества израсходованного живого и овеществленного труда, выраженного, например, в затратах C , в результат E по формуле

$$\dot{u}_{ek} = \frac{C}{E} = \max. \quad (2.7)$$

Акции и реакции, выраженные в технических единицах, дают характеристику **технической эффективности** $\dot{u}_t$ по которой изучается величина требуемых усилий P , необходимых для достижения ожидаемого результата V , по формуле

$$\dot{u}_t = \frac{V}{P} = \max. \quad (2.8)$$

Целесообразность объекта (процесса) управления выступает в форме усилий, прилагаемых для достижения **оптимальных** результатов.

Если в соотношении (2.8), выражающем техническую эффективность, заменить категорию усилий P **экономической** категорией, выражающей затра

ты на функции объекта (процесса) $\sum_{i=1}^n C_{ij}$, а категорию результата V поменять

на качественное понятие **функциональности** в виде степени выполнения

функций объекта (системы) $\sum_{i=1}^n {}^0F_{ij}$, то получим текущую (нынешнюю) форму

критерия эффективности в функционально-стоимостном анализе. Им является **показатель относительной эффективной стоимости** $S_{отэфj}$, который получается как соотношение

$$S_{отэфj} = \frac{\sum_{i=1}^n {}^0F_{ij}}{\sum_{i=1}^n C_{ij}} = \max, \quad (2.9)$$

где i — функция ($i = 1, 2, 3, \dots, n$);

j — вариант ($j = 0, 1, 2, \dots, m$);

$({}^0)$ — означает текущее (нынешнее) состояние управления объектом (системой).

Показатель относительной эффективной стоимости отражает комплексную эффективность в условиях производственных отношений. Комплексная эффективность в функционально-стоимостном анализе понимается как последовательно соблюдаемое отношение функциональности к затратам. В процессе стоимостного анализа необходимо всегда стремиться достичь оптимальной функциональности или оптимального удовлетворения общественных потребностей данным объектом (системой). Величина производимых затрат рассматривается по отношению к уровню функциональности. Однако необходимо отметить постоянную тенденцию к минимизации затрат труда, необходимых для обеспечения **оптимального**, то есть общественно необходимого выполнения функций [2,45].

Целью каждого случая применения функционально-стоимостного анализа является максимизация соотношения трансформируемой потребительной

$$\sum_{i=1}^n {}^0F_{ij} \quad \sum_{i=1}^n C_{ij} .$$

стоимости и затрат на ее
обеспечение

Соотношение этих величин может изменяться в следующих основных направлениях.

Равномерное движение означает одинаковый темп и пропорциональность расчета обеих величин, входящих в показатель относительной эффективной стоимости. Для предприятия эффективным будет такое решение, которое обеспечит большую рентабельность по сравнению с базовым объектом (системой). Такое управленческое решение принесет и более высокий эффект потребителю (заказчику). Однако при этом показатель относительной эффективной стоимости может остаться неизменным, если эти решения касаются потребительных стоимостей различного качества. (Этот недостаток обуславливается природой самого показателя, основанного на соотношении разнохарактерных величин).

Неравномерное движение величин показателя $S_{отэф,j}$ характеризует степень выполнения функций и затраты на их обеспечение, динамика которых не сохраняет прямой пропорциональности.

В соответствии с целью стоимостного анализа должен расти показатель относительной эффективной стоимости, а вместе с ним и общая эффективность проектируемого решения

$S_{отэф,1}$ по сравнению с исходными данными

$S_{отэф,0}$

Желаемый рост **функциональности** достигается при уменьшении темпов роста затрат. Этот случай, соответствующий зоне 1 (рис.2.2), можно с помощью символов одной функции выразить следующим образом

$${}^0F_{i1} > {}^0F_{i0}, \quad C_{i1} > C_{i0}$$

при условии, что прирост 0F_1 больше прироста C'_{i1} .

$$[\text{Прирост (уменьшение) здесь есть: } \frac{{}^0F_{i1}}{{}^0F_{i0}} \text{ либо } \frac{{}^0F_{i1} - {}^0F_{i0}}{{}^0F_{i0}}, \\ \frac{C_{i1}}{C_{i0}} \text{ либо } \frac{C_{i1} - C_{i0}}{C_{i0}}].$$

Желаемый рост функциональности достигается при неизменных затратах, что характеризует вектор в зоне 2. С помощью символов одной функции это можно выразить следующим образом

$${}^0F_{i1} > {}^0F_{i0}, C_{i1} = C_{i0}. \quad (2.11)$$

Желаемый рост функциональности достигается при уменьшении затрат. Этот вариант решения графически характеризуется множеством точек в зоне 3. С помощью символов одной функции это выражается так

$${}^0F_{i1} > {}^0F_{i0}, \text{ а } C_{i1} < C_{i0}. \quad (2.12)$$

Желаемое сохранение функциональности при уменьшении затрат графически иллюстрируется множеством точек в зоне 4, а с помощью обозначения одной функции выражается так

$${}^0F_{i1} = {}^0F_{i0}, \text{ а } C_{i1} < C_{i0}. \quad (2.13)$$

Желаемое уменьшение функциональности до уровня **оптимально достигаемого** при быстром снижении затрат графически характеризуется множеством точек в зоне 5, а в математической интерпретации с помощью символов одной функции выглядит следующим образом

$${}^0F_{i1} < {}^0F_{i0}, \text{ а } C_{i1} < C_{i0}, \quad (2.14)$$

уменьшение ${}^0F_{i1}$ меньше уменьшения C_{i1} . Из приведенной выше характеристики критерия эффективности в функционально-стоимостном анализе следует, что относительная эффективная стоимость является показателем **уровня выгодности** или ее определенной мерой.

Этот показатель можно также охарактеризовать как меру или коэффициент эффективности [2].

При оценке некоторых специфических проблем технического характера можно для большей наглядности использовать различные модификации показателя относительной эффективной стоимости. Вместо затрат либо цены применяется абсолютная или относительная стоимость, выражающая материальные или иные (качественные) величины. Например, в авиации, ракетной технике, космических исследованиях - это материалоемкость приборов и оборудования, стоимость которых влияет на степень выполнения функций. В других случаях, например, в производстве различного оборудования, обеспечивающего безопасность работы человека, целесообразнее использовать стоимость, характеризующую **величину риска**. Можно, однако, эти величины выразить де-

нежными единицами и в затратах либо в цене охарактеризовать количество труда, необходимого для минимизации **материалоемкости** или **риска**.

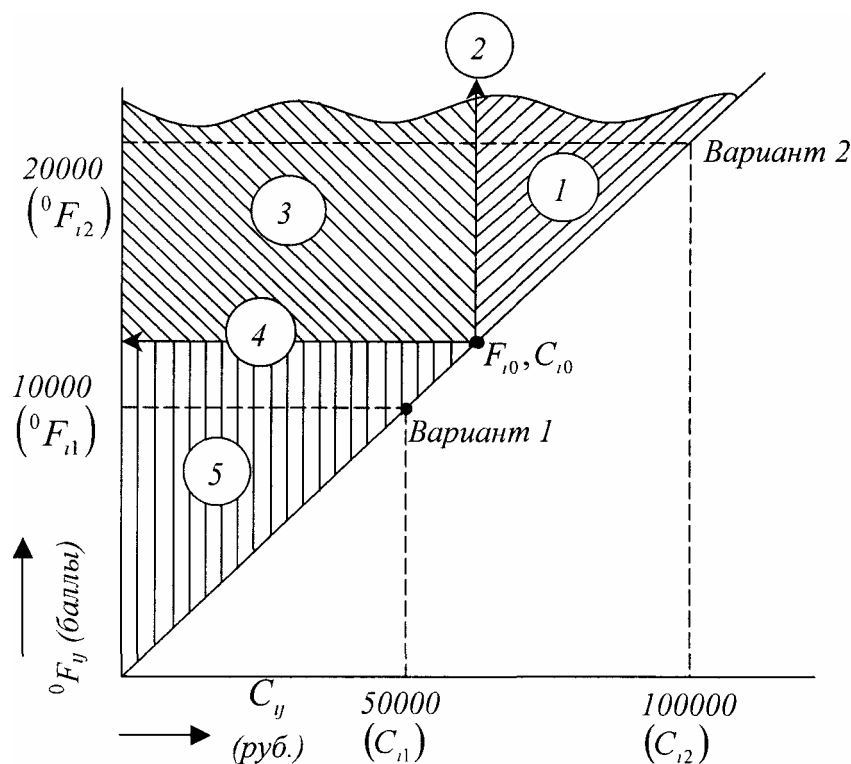


Рис. 2.2. Варианты эффективности, выраженные показателем относительной эффективной стоимости

В практике стоимостного анализа долгое время применялся только один способ расчета показателя относительной эффективной стоимости, который не всегда был точным.

Неточность проявлялась в том, что в числителе оценивалась степень выполнения другого комплекса функций, а не того, затраты на который были учтены. Очень часто случается, что в числителе этого показателя отражается степень выполнения структурных, потребительских и обслуживающих функций оцениваемого изделия, а в знаменателе приводится только полная себестоимость объекта анализа у производителя. Деление функций с вещественной точки зрения создает предпосылки для более точного выражения эффективности в функционально-стоимостном анализе.

В зависимости от сферы и области применения функционально-стоимостного анализа определяют следующие показатели относительной эффективной стоимости управления:

1. Эффективность проектируемых функций управления - S_1 . Показатель выражает эффективность подготовки нового объекта (изделия, процесса, системы). При его построении исходят из следующих данных: **степени вы-**

полнения проектируемых функций $\Sigma^{\circ}F_I$, затрат на исследования, развитие и техническую (технологическую) подготовку производства ΣC_I ;

2. Эффективность производственных функций - S_{II} . Показатель характеризует эффективность способа производства продукции. При его составлении исходят из следующих данных: **степени выполнения производственных функций** $\Sigma^{\circ}F_{II}$, затрат на производство, то есть полной себестоимости за минусом материальных затрат ΣC_{II} ;

3. Эффективность структурных функций - S_{III} . Показатель выражает эффективность производимой продукции с позиции производителя (предприятия-подрядчика). При его составлении применяются следующие данные: степень выполнения структурных функций продукта $\Sigma^{\circ}F_{III}$, полная себестоимость продукции ΣC_{III} ;

4. Эффективность первичных функций - S_{IV} . Показатель выражает народно-хозяйственную эффективность продукции. В нем учтена эффективность как производителя, так и потребителя, а также **общественные критерии оценки продукции**. При построении показателя исходят из следующих данных: степени выполнения всех первичных функций $\Sigma^{\circ}F_{IV}$, общественных затрат на оцениваемый продукт (объект, изделие, систему) $\Sigma^{\circ}F_{IV}$.

Последующая составляющая подразумевает затраты с различным временным фактором, то есть единовременные I и годовые эксплуатационные затраты P . Единовременные затраты состоят из расходов на организацию, включая затраты на приобретение лицензий, проведение исследований, техническое развитие и затраты по реализации продукции (объектов, систем, изделий, т.д.). Полные годовые эксплуатационные затраты включают расходы, связанные с собственно эксплуатацией, ремонтом, содержанием и обслуживанием изделия. Для устранения проблемы, связанной с равномерностью затрат I и P , проводится их корректировка путем приведения к показателю приведенных затрат

$$\sum C_{IV} = I \cdot k_n + P, \quad (2.15)$$

где k_n - коэффициент перевода единовременных затрат, сконструированный как обратная величина срока окупаемости.

Способы оценки эффективности управленческой деятельности можно разделить на две группы:

1. Оценка эффективности управленческой деятельности в узком смысле;
2. Оценка эффективности управленческой деятельности в широком смысле.

При **оценке эффективности управленческой деятельности в узком смысле** оцениваются только результаты деятельности самого субъекта управления. Понимаемая таким образом эффективность управленческой деятельности характеризуется, например:

1. **Удельным весом** работников, занятых в управленческом аппарате, по всем работающим в данном производственном коллективе (организации, заводе, объединении и т.д.);

2. **Уровнем механизации и автоматизации** управленческой деятельности (например, показателем технической оснащенности работы одного управленца, то есть рабочего места одного работника управленческого аппарата);

3. Структурой использования рабочего времени управленческим аппаратом;

4. Численностью работников управленческого аппарата.

Сравнение приведенных показателей функционально-стоимостного анализа свидетельствует о значительной проблематичности использования того или иного способа выражения эффективности управленческого труда.

Несколько лучшую, хотя тоже проблематичную возможность выражения эффективности управленческой деятельности отражают показатели, характеризующие действенность собственно управленческого процесса. Здесь речь идет о применении такого качественного показателя, который дает представление о количестве информации, созданной самими работниками аппарата управления. Это может быть **показатель производительности управленческой деятельности**, опосредованно отражающий ее эффективность. Количество информации может быть выражено в виде числа решений, докладов, уведомлений и т.д., приходящихся на одного работника управленческого аппарата в единицу времени. Конечно, эти данные ничего не говорят о качестве управленческой деятельности. В связи с этим возникает вопрос, можно ли устранить такой серьезный недостаток введения качественной характеристики информации, с помощью единицы измерения, которая в литературе по автоматизации (системотехнике) обозначается, как один **бит**.

При оценке **эффективности управленческой деятельности в широком смысле** принимается во внимание влияние управленческой деятельности на конечные результаты, достигаемые управляемым объектом (изделием, системой, коллективом и т.п.). Такой подход, учитывающий вклад управленческой деятельности в достижение конечных результатов, представляется более логичным по сравнению с любым другим. Например, на авиационном (равно как и на любом промышленном) предприятии эффективность управленческой деятельности могла бы быть выражена с помощью соответствующим образом подобранных «стоимостей» выполнения обязательных показателей, какими, например, являются рентабельность производственных фондов (производственных факторов), рост производительности труда, цена единицы выбранных экспортных товаров, удельный вес изделий высшей категории качества, удельный вес новых изделий (новинок), современных изделий и изделий, например, серии «люкс», поставки на экспорт по двум направлениям - в развитые капиталистические страны и страны третьего мира (то есть развивающиеся), поставки

на отечественный рынок для производственного потребления, транспортного сообщения (самолеты, космические аппараты и т.п.), достижение обязательных технико-экономических параметров, соблюдение сроков реализации заданий научно-технического развития (прогресса) и т.д. Свою роль в определении эффективности управленческой деятельности могли бы сыграть и другие измеряемые количественно показатели, характеризующие усилия управленческой сферы в определенных областях, например, удельный вес внутрипроизводственного брака, рекламаций заказчиков, количество штрафов внутри отрасли, комплекса (или между поставщиками - смежниками) и др.

Упорядочение этих соответствующим образом подобранных показателей для определения эффективности управленческой деятельности заключается в преобразовании их в форму коэффициентов. Количественные данные, выражающие общую эффективность управленческого труда, образуют коэффициент, **называемый коэффициентом эффективности управления**. Он рассчитывается как произведение всех коэффициентов избранных показателей, которые отражают степень выполнения оцениваемых ими явлений. При расчете коэффициента эффективности управления необходимо, однако, принимать во внимание также и значения отдельных показателей и их коэффициентов.

Основным критерием для определения выбираемых показателей могут быть главные задания организации в оцениваемом периоде.

Несмотря на определенную спорность конструкции коэффициента эффективности управления понятно, что его использование является одним из реальных способов выражения эффективности управленческой деятельности. Чем ближе будет коэффициент эффективности управления приближаться к единице, тем выше будет эффективность управленческой деятельности на оцениваемом объекте (предприятии, коллективе и т.п.).

Здесь рассматривается только одна сторона эффективности процесса управления - это его **полезность**. Поэтому следует уточнить термин «эффективность» для данного случая. **Эффективность**, правильно понимаемая как **действенность и целесообразность** оцениваемого явления в комплексном понимании должна быть выражена как отношение «выпуска», то есть пользы управленческого процесса, к «затратам» на этот управленческий процесс. В сфере управления при принятии оптимальных решений и проведении их в жизнь важную роль играет время. Поэтому при определении эффективности управления необходимо обязательно учитывать фактор времени.

При комплексном понимании эффективности для ее измерения может быть применен известный показатель **относительной эффективной стоимости** в следующем виде

$$S_{om \text{ эф см}} = \frac{\sum_{i=1}^{n_i} {}^0 F_y}{\sum_{i=1}^n C_y \cdot t} = max, \quad (2.16)$$

где $\sum_{i=1}^{n_i} F_{ij}$ - функциональность эффективности управления;

$\sum_{i=1}^n C_{ij}$ - затраты на подготовку и реализацию решений;

t - время на подготовку и реализацию решений.

Применение этого показателя отражает понимание того, что эффективность управления есть ничто иное, как достижение поставленной цели данными средствами в установленное время.

2.3. Сравнительная характеристика функционального синтеза оптимального управления процессами и системами

Под **функциональным синтезом** оптимизации управления понимается сложный процесс творческого мышления и критических оценок, целью которого является поиск эффективного способа осуществления всех **требуемых функций**. Эта деятельность направлена на постепенное отыскание эффективных способов обеспечения прежде всего **главной**, а потом и **вспомогательных** функций. При этом сначала однозначно формируются функции, а затем начинается разработка максимального числа предложений по их обеспечению. В дальнейшем предложения преобразовываются в реальные варианты, то есть проекты обеспечения всех функций. Функциональный синтез заканчивается выбором и формулировкой проекта оптимального варианта объекта. В качестве критерия **оптимизации** для выбора самого эффективного решения применяется показатель относительной эффективной стоимости.

Функциональный синтез оптимального управления начинается еще в фазе функционального анализа. Элементом функционального синтеза управления нужно считать прежде всего процесс упорядочения функций управления. Речь идет об определении **главных и вспомогательных** функций в виде логического дерева функций либо их графика. Он применяется для упорядочения функций таких объектов, которые имеют характер процесса.

В рамках функционального синтеза применяется ряд элементов функционального анализа. Это, в частности, уточнение формулировок функций перед применением некоторых методов творческого мышления и прежде всего инструментов оценки функций, отдельных вариантов решений. Функциональный анализ и функциональный синтез - это взаимосвязанные и органически переплетающиеся процессы реализации функционального принципа, применяемого при создании либо совершенствовании избранного объекта. Здесь подразумевается и «реализуется» конкретное проявление **диалектического единства** процесса анализа и синтеза в творческой человеческой деятельности, основанной на функциональном принципе. Сущность функционального синтеза в связи с функциональным анализом отражает рисунок 2.3.

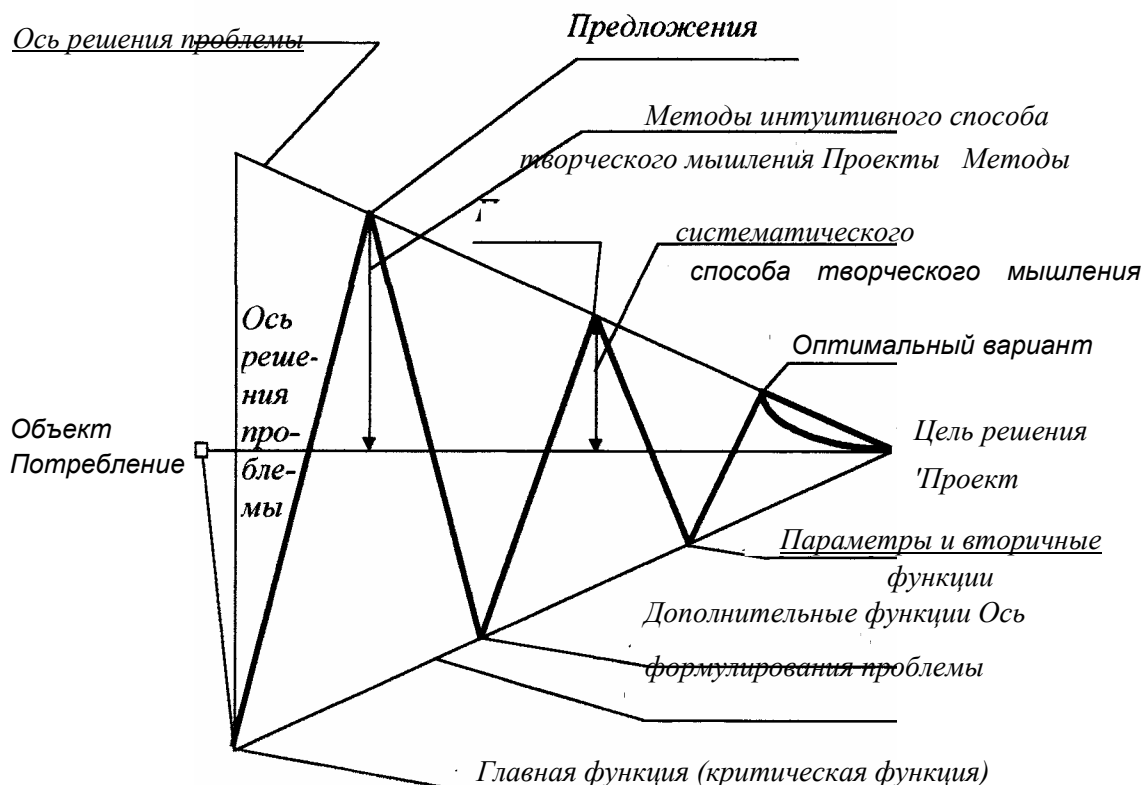


Рис.2.3. Схема процесса функционального синтеза системы управления

Из приведенной характеристики функционального синтеза вытекает, что сущностью является органическое единство процесса создания идей и проектов новых вариантов решений и их критической оценки. Этот творческий процесс реализуется в большинстве случаев в условиях коллективной работы и при использовании ряда методов, которые можно разделить на большие группы. Это методы **творческого мышления** и методы **оценки вариантов**.

2.4. Процедуры функционально-стоимостного анализа для оценки оптимального управления процессами и системами

Под управлением (в смысле функционально - стоимостного анализа) понимается специфическое воздействие одного элемента (управляющего субъекта) на другой элемент (управляемый объект) с целью перевода управляемого объекта из одного состояния в другое. В ходе управления производственными, равно как и общественными процессами, а в них - экономическими системами процесс управления протекает как специфическая деятельность людей. Субъектом и объектом управления являются люди и при их посредничестве экономические системы. Речь идет о таких системах, в рамках которых протекают хозяйственные процессы, представляющие материальные процессы производства, распределения, обмена и потребления потребительных стоимостей.

Управление хозяйственными процессами является поэтому общественно значимой и целенаправленной деятельностью [2, 3,4].

Процессы управления экономическими системами являются в целом и в частности удобными объектами для применения функционально-стоимостного анализа в целях оптимизации управленческих процессов в разных его формах и модификациях.

Исходя из теоретических представлений о структуре управления, механизм процесса управления следует понимать как неразрывное единство актов принятия решений, проведения этих решений в материальную сферу человеческой деятельности и контроля за их реализацией. Все эти звенья механизма управления связаны между собой в общем информационном процессе.

Под процессом принятия решения понимается выработка решения управляющим субъектом о развитии управляемого объекта, причем основой принятия решения является выбор оптимального варианта дальнейшего развития управляемого объекта. Конечным результатом принятия решения выступает решение управляющего субъекта в форме информации, устраняющей неопределенность в предстоящем развитии управляемого объекта.

Под процессом **проведения решения в жизнь** понимается перенесение информации (решения), оптимальной по наполнению и смыслу, от управляющего субъекта к управляемому объекту. Этим перенесением решения оказывается воздействие на проведение управляемого объекта как наполнителя решений управляющего субъекта.

Под процессом **оптимального контроля** понимается воздействие обратной связи как на процесс принятия решений, так и на их реализацию. Смысл процесса контроля заключается в выявление отклонений в реальном развитии управляемого объекта от заданного, определяемого решениями. Контроль осуществляется с целью устранения этих отклонений либо изменением первоначальных решений.

Три звена механизма управления, то есть процессы принятия решений, реализации и контроля, создают так называемый процесс управления в **узком смысле**. Процессы принятия решений, реализации и контроля однако всегда протекают как комплексное действие управляющего субъекта на управляемый объект с целью перевода этого объекта в другое состояние, ограничения либо полного устранения неопределенности в его предстоящем развитии.

Перевести управляемый объект из одного состояния в другое, ограничить либо устранить неопределенность его будущего развития означает направить управленческую деятельность к трем целям, то есть:

1. Организовать управление объектом так, чтобы он существовал как хозяйственная (производственная) система;
2. Управлять конкретными процессами;

3. Управлять развитием личности работника и трудового коллектива как носителями и реализаторами процессов, протекающих в организованных структурах хозяйственных (производственных) систем.

Общий процесс оптимального управления или **управления в широком смысле** возникает в результате интеграции этих трех видов деятельности, то есть организации управления, управления конкретными процессами и управления развитием личности. Механизм управления для всех этих видов управленческой деятельности по существу одинаковый. Он обеспечивается единством процессов принятия решений, реализации и контроля.

Анализ структуры процесса управления хозяйственными (производственными) системами, характеристика отдельных звеньев механизма управления и видов процесса управления вместе с характеристикой функционально-стоимостного анализа и накопленным практическим опытом являются исходным материалом для формулирования концепций применения функционально-стоимостного анализа в системе управления и оптимизации управленческих решений.

Цель любого процесса принятия решения - выработать своевременно такое решение, которое направило бы управляемый объект к достижению более высокой эффективности хозяйственной (производственной) системы. Рассмотрение вариантов оптимальных решений, отбор одного из них, наиболее оптимального, и формулирование окончательного решения представляют собой логически связанный и целенаправленный процесс в деятельности управляющего субъекта. В специальной литературе предлагаются различные классификации отдельных фаз процесса принятия оптимального решения. Однако все они похожи друг на друга в основных чертах, по сущности, по целям процесса принятия решения. Изучение содержания отдельных фаз процесса принятия решения позволяет сформулировать методологические принципы применения функционально-стоимостного анализа. При решении относительно сложных проблем можно встретить следующие фазы:

1. Выявление проблемы. На основе изучения анализа информации делается заключение о существовании проблемы;
2. Принятие решения о разработке проблемы. Принимается решение: будет или не будет решаться проблема;
3. Анализ проблемы. Проблема изучается со стороны содержания, времени, методологии, оцениваются имеющиеся в распоряжении источники и средства, определяются критерии оценки, компетенция, правомочия;
4. Прогноз будущего положения. Высказываются предложения о предстоящем развитии решаемой проблемы, определяется величина риска, связанного с возможными направлениями развития;

5. Разработка вариантов решения. Осуществляются поиск и разработка вариантов возможного решения проблемы. Если нет хотя бы двух вариантов решения, нельзя говорить о решении проблемы;

6. Контроль предполагаемых вариантов решений. По определенным критериям оцениваются варианты, их соответствие основным целям, потребностям и интересам управляющего субъекта;

7. Выбор определенного варианта решения. В качестве наилучшего выбирается тот вариант, который более других соответствует критериям оценки. С информационной точки зрения решение является трансформацией полученной информации в новую информацию (управляющую информацию);

8. Формулирование решения. Эта фаза представляет переход от информационной подготовки решения к его реализации. Речь идет о формальной подготовке решения как проявлении воли управляющего субъекта, который эту волю выражает в удобной форме, вразумительно и своевременно.

Сравнение методологии функционально-стоимостного анализа с содержанием приведенных фаз процесса принятия решения не оставляет никакого сомнения в том, что речь идет по существу о методологических подходах, которые по своим целям, последовательности проведения весьма близки, если не тождественны между собой. Этот факт говорит о том, что есть все возможности для систематического внедрения функционально-стоимостного анализа в качестве целенаправленного методологического инструмента в процесс разработки **управляющих воздействий и принятия решений**. Можно также сказать несколько по другому так: **стоимостной анализ** является инструментом решения проблемы, этапы его идентичны фазам процесса принятия решений. Этот вывод создает основу для постоянного внедрения стоимостного анализа в этом звене механизма управления. Соотношение фаз процесса принятия решения с этапами стоимостного анализа характеризует табл.2.1.

Специфика процесса принятия решений [2] находит отражение в методологии функционально-стоимостного анализа. Речь идет прежде всего о проблеме риска, связанного с принятием решений в рамках управления общественными процессами, а в них - управления хозяйственными системами.

С точки зрения риска, связанного с принятием решений, различаются три основных типа решений:

1. Решения бесспорные;
2. Решения с возможностью риска;
3. Решения спорные.

Таблица 2.1

**Фазовое пространство процессов принятия решения и этапов
стоимостного анализа**

Фазы процесса принятия решения	Этапы функционально-стоимостного анализа
1. Выявление проблемы.	I. Выбор объекта.
2. Принятие решения о разработке проблемы.	II. Сбор информации.
3. Анализ проблемы.	III. Функциональный анализ.
4. Прогноз будущего положения.	IV. Разработка проекта обеспечения функций.
5. Разработка вариантов решения.	V. Оценка проекта.
6. Контроль предлагаемых вариантов решения.	VI. Разработка проекта оптимального варианта объекта.
7. Выбор определенного варианта решения.	VII. Обсуждение и утверждение проекта.
8. Формулирование решения.	VIII. Выполнение проекта.

Модификация функционально-стоимостного анализа вследствие существования риска в ситуациях, требующих решений, вызывается необходимостью учета способа определения пользы или расчета уровня либо степени выполнения функций в рамках отдельных вариантов объекта анализа. Эта корректировка методологического подхода при определении уровня выполнения функций осуществляется только в случаях принятия решений с возможностью риска либо спорных решений [2, 3, 5].

При принятии решений с возможностью риска учитывается польза либо степень выполнения функций, которые могут быть достигнуты в условиях определенного положения (5,) либо ситуации, складывающейся с определенной

вероятностью (P_i) , причем $\sum_{i=1}^n P_i = 1$. В этом случае учитывается так называемая средняя величина ожидаемой степени выполнения функций $({}^0\dot{F}_y)$ как сумма слагаемых реальной степени выполнения функций и вероятности возникновения ситуации в соответствии с формулой

$${}^0\dot{F}_y = \sum {}^0F_y P_i. \quad (2.17)$$

Если исходить из степени выполнения функций, то оптимальным может быть тот вариант, у которого средняя величина ожидаемой степени выполнения функций будет самой высокой (табл. 2.2).

$${}^0F'_A = 0,2 \cdot 7 + 0,2 \cdot 8 + 0,5 \cdot 5 + 0,1 \cdot 9 = 6,4$$

$${}^0F'_B = 0,2 \cdot 10 + 0,2 \cdot 12 + 0,5 \cdot 9 + 0,1 \cdot 5 = 9,4$$

$${}^0F'_C = 0,2 \cdot 8 + 0,2 \cdot 10 + 0,5 \cdot 10 + 0,1 \cdot 9 = 8,9$$

Таблица 2.2

Средняя величина степени выполнения функций

Варианты объекта анализа	Состояние среды	S_1	S_2	S_3	S_4
	Вероятность (P_I)	0,2	0,2	0,5	0,1
A	$^oF_Y'$	7	8	5	9
B		10	12	9	5
C		8	7	10	9

Оптимальным является вариант **B**, поскольку средняя величина ожидаемой степени выполнения функций у него является наибольшей ($^oF_B = 9,4$).

При принятии спорных решений выбор вариантов затруднен тем, что вероятность (P_I), с которой складывается положение или определенная ситуация

(S_1), **неизвестна**.

В этих случаях возможно следующее:

1. Определяем величину вероятности (P_I) и тем самым принятие спорного решения переводим в ситуацию принятия решения в условиях риска;

2. Применяем для целей функционально-стоимостного анализа:

правило Лапласа. Оно предполагает, что все будущие состояния объекта анализа одинаково вероятны. **Оптимальным** является тот вариант объекта анализа (управления), при котором данная величина ожидаемой степени выполнения функций является наибольшей;

правило максиминимума (maximin). У каждого варианта выбирается самая низкая величина степени выполнения функций ($^oF_b = \min$). Оптимальный вариант тот, у которого наибольшая величина минимальной степени выполнения функций;

правило максимаксима (maximax). У каждого варианта выбирается самое большое значение степени выполнения функций, а потом избирается та альтернатива, которая максимизирует максимум степени выполнения функций.

Что касается приспособления функционально-стоимостного анализа к особенностям процесса принятия решений при управлении хозяйственными процессами, то следует прежде всего охарактеризовать вклад указанного анализа в этот процесс. При принятии решений величина будущей пользы от их реализации рассматривается как критерий, исходя из которого и принимаются решения. Трудовые затраты для достижения этой пользы могут, но не обязательно должны быть частью этого критерия.

При функционально-стоимостном анализе в качестве критерия эффективности выступает отношение достигнутой пользы к количеству затраченного труда, то есть к затратам, необходимым для достижения этой пользы. Критерием эффективности служит показатель относительной стоимости.

Теория и практика функционально-стоимостного анализа подтверждает пользу (полезность) применения этого метода на всех стадиях (звеньях) процесса в организации управления, управлении конкретными процессами и управлении развитием личности и трудовых коллективов. Однако для отдельных видов управленческой деятельности необходимо применение специфического методологического подхода, что предопределяет дифференцированный подход к проведению функционально-стоимостного анализа этих объектов.

Таким образом, **оптимизация управления** с позиции функционально-стоимостного анализа составной части функционально-стоимостной инженерии на современном этапе является весьма уместным мероприятием, полезным с позиции **функциональности и стоимости**, так как **функционально-стоимостная инженерия** - это комплекс методов с технико-экономическим, математическим, электронным или натурным априорным моделированием структуры и состава управляемого объекта (изделия, системы и т.п.), сутью которого является поиск и нахождение лучшего или даже совершенно нового решения функций вновь проектируемого или модернизируемого изделия (объекта, системы и т.п.) с целью обеспечения его **функциональности** (т.е. качества, надежности, эффективности, ремонтпригодности и долговечности), **стоимости** и реальной конкурентоспособности.

2.5. Анализ оптимального управления на основе марковских процессов принятия решения

Пусть при любом начальном состоянии и принятом в нем решении процесс перешел в некоторое новое состояние. Тогда если исходная стратегия управления была **оптимальной**, то и ее оставшаяся часть тоже оптимальна для процесса, начинающегося из нового состояния.

Принцип **оптимальности** принятия управленческих решений был впервые сформулирован Беллманом. Используя этот принцип, получаем рекуррентное соотношение

$$V_{n+1}^i(\Pi^*) = \max_{k \in K_i} \left[r_i^k + \sum_{j \in S} P_{ij}^k V_n^j(\Pi^*) \right], \quad (2.18)$$

справедливое для всех $i \in S$ и любого $n \geq 0$,
где $V_0^i(\Pi^*) = 0$ для всех $i \in S$.

(2.19)

Соотношения (2.18), (2.19) позволяют находить оптимальную стратегию управления.

Отметим, что здесь применяется метод динамического программирования, который применим и в случае, когда величины r_i^k и P_y^k зависят от n , при этом рекуррентные соотношения (2.18) и (2.19) остаются справедливыми.

Итак, стратегия управления Π называется **оптимальной**, если для любой стратегии управления σ и любого $n \geq 1$ выполняется неравенство $V_n(\Pi) \geq V_n(\sigma)$ [6].

Теперь займемся изучением оптимальных стратегий управления для процессов с **бесконечным временем планирования**. Предложим, что величины r_i^k и P_y^k не зависят от n . Воспользуемся оператором L , переводящим $(N \times 1)$ -мерный вектор X в вектор $L(f)X = r(f) + P(f)X$ той же размерности. Для любой стратегии Π и любого $(N \times 1)$ -мерного вектора определим

$$V^n(\Pi, X) = L(f_n)L(f_{n-1})\dots L(f_1)X = r(f_n) + P(f_n)r(f_{n-1}) + \dots + P(f_n)\dots P(f_2)r(f_1) + P(f_n)\dots P(f_1)X. \quad (2.20)$$

Вектор $V^n(\Pi, X)$ можно интерпретировать как вектор суммарных средних доходов, получаемых за n шагов, при условии, что в момент окончания последнего перехода выплачивается сумма, равная j -й компоненте вектора X , если при этом процесс попадает в состояние j . В частности, получаем, что $V^n(\Pi, 0) = v_n(\Pi)$.

Следовательно, формулируем **определение**: стратегия управления Π называется **X -оптимальной**, если для любой стратегии σ и любого $n \geq 1$ выполняется неравенство

$$V^n(\Pi, X) \geq V^n(\sigma, X).$$

Из приведенного определения следует, что стратегия Π является оптимальной тогда и только тогда, когда она O -оптимальна [14].

Лемма 2.5.1. Для любого элемента $f \in F$ имеет место равенство

$$V_\beta(f^\infty) = \sum_{i=0}^{\infty} \beta^i P^i(f)r(f) = [(1-\beta)^{-1} - 1]u(f) + A(f) + E(\beta, f), \quad (2.21)$$

где $E(\beta, f) \rightarrow 0$ при $\beta \rightarrow 1-0$, а $A(f) = [I - P(f) + P^*(f)]^{-1}r(f)$, согласно следующей теореме.

Теорема 2.5.1.1. Для любой стационарной стратегии управления f^∞ по-

ложим
$$P^*(f) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} P^i(f), \quad (2.22)$$

то есть пусть P – стохастическая матрица размера $N \times N$. Тогда:

1. Последовательность $\sum_{i=0}^{n-1} P^i / n$ сходится при $n \rightarrow \infty$ к стохастической матрице P^* такой, что

$$PP^* = P^*P = P^*P^* = P^*; \quad (2.23)$$

2. Матрица $I - (P - P^*)$ невырождена и

$$H(\beta) = \sum_{i=0}^{\infty} \beta_i (P^i - P^*) \rightarrow H = (I - P + P^*)^{-1} - P^* \quad (2.24)$$

при $\beta \rightarrow 1 - 0$,

$$H(\beta)P^* = P^*H(\beta) = HP^* = P^*H = 0 \quad (2.25)$$

и

$$(I - P)H = H(I - P) = I - P^*; \quad (2.26)$$

$$3. \text{ Ранг}(I - P) + \text{ранг } P^* = N; \quad (2.27)$$

4. Для любого $(N \times 1)$ -мерный вектор c

$$Px = x, P^*x = P^*c \quad (2.28)$$

имеет единственное решение x .

$$\text{Тогда } V_{\beta}(f^{\infty}) = u(f)/(1 - \beta) + v(f) + \varepsilon(\beta, f), \quad (2.29)$$

где $u(f)$ – естественное решение уравнений

$$[I - P(f)]u = 0, P^*(f)u = P^*(f)r(f), \quad (2.30)$$

$v(f)$ – единственное решение уравнений

$$[I - P(f)]v = r(f) - u(f), P^*(f)v = 0, \quad (2.31)$$

а $\varepsilon(\beta, f) \rightarrow 0$ при $\beta \rightarrow 1 - 0$.

Докажем эту теорему (2.5.1.1.).

$$\begin{aligned} \text{Имеем } V_{\beta}(f^{\infty}) &= \sum_{i=0}^{\infty} \beta_i P^i(f)r(f) = \left\{ \sum_{i=0}^{\infty} \beta_i [P^*(f) + \sum_{i=0}^{\infty} \beta^i [P^i(f) - P^*(f)]] \right\} r(f) = \\ &= \frac{P^*(f)r(f)}{1 - \beta} + H(f)r(f) + [H(\beta, f) - H(f)]r(f), \end{aligned} \quad (2.32)$$

где $H(f)$ и $H(\beta, f)$ определяются по формуле

$$H(\beta) = \sum_{i=0}^{\infty} \beta_i (P^i - P^*) \rightarrow H = (I - P + P^*)^{-1} - P^* \quad (2.33)$$

при $\beta \rightarrow 1 - 0$. Таким образом, равенство (2.29) получится, если положить

$$u(f) = P^*(f)r(f), v(f) = H(f)r(f) \quad \text{и} \quad \varepsilon(\beta, f) = [H(\beta, f) - H(f)]r(f).$$

Совершенно очевидно из (2.23) по (2.28) следует, что $u(f)$ и $v(f)$ являются единственными решениями (2.30) и (2.31) соответственно.

Решения $u(f)$ и $v(f)$ уравнений (2.30) и (2.31) могут быть представлены в виде

$$u(f) = P^*(f)r(f), \quad (2.34)$$

$$v(f) = H(f)r(f) = \left\{ [I - P(f) + P^*(f)]^{-1} - P^*(f) \right\} r(f). \quad (2.35)$$

Таким образом, задача заключается в нахождении стратегий управления, максимизирующей $u(f)$, а также $u(f)$ на множестве всех стационарных стра-

тегий управления. Для ее решения необходим эффективный алгоритм, чтобы задачу можно было бы поставить на ЭВМ.

Существует стационарная оптимальная стратегия управления по критерию среднего дохода.

Определим множество политик

$$\dot{F} = \{f / f \in F, u(f) \geq u(g) \text{ для всех } g \in F\}. \quad (2.36)$$

Иначе говоря, $\dot{F}$ – множество всех $f \in F$, при которых достигается максимальный средний доход за единицу времени. Далее, определим множество политик

$$F'' = \{f / f \in \dot{F}, v(f) \geq v(g) \text{ для всех } g \in \dot{F}\}. \quad (2.37)$$

Очевидно, что $\ddot{F}$ – подмножество $\dot{F}$ и является множеством всех $f \in \dot{F}$, максимизирующих член $v(f)$, характеризующий смещение.

Далее с переоценкой вектор $V_\beta(g, f^\infty)$ согласно (2.29) равен

$$\begin{aligned} V_\beta(g, f^\infty) &= r(g) + \beta P(g) V_\beta(f^\infty) = \beta P(g) u(f) / (1 - \beta) + r(g) + \\ &+ \beta P(g) v(f) + \beta P(g) \varepsilon(\beta, f) = P(g) u(f) / (1 - \beta) + r(g) - P(g) u(f) + \\ &+ P(g) v(f) - (1 - \beta) P(g) v(f) + \beta P(g) \varepsilon(\beta, f) = P(g) u(f) / (1 - \beta) + \\ &+ r(g) - P(g) u(f) + P(g) v(f) + \varepsilon_1(\beta, f, g), \end{aligned} \quad (2.38)$$

где $\varepsilon_1(\beta, f, g) = -(1 - \beta) P(g) v(f) + \beta P(g) \varepsilon(\beta, f) \rightarrow 0$ при $\beta \rightarrow 1 - 0$.

Сравнивая правые части равенств (2.29) и (2.38), определим множество решений

$$\begin{aligned} G(i, f) &= \{k / k \in K_i, \sum_{j \in S} P_{ij}^k u_j > u_i \text{ или} \\ &\sum_{j \in S} P_{ij}^k u_j = u_i \text{ и } r_i^k + \sum_{j \in S} P_{ij}^k v_j > u_i + v_i\}, \end{aligned} \quad (2.39)$$

где u_i и v_i – i -е компоненты векторов $u(f)$ и $v(f)$ соответственно.

Определим далее

$$G(f) = G(1, f) \times G(2, f) \times \dots \times G(N, f). \quad (2.40)$$

Таким образом, рассмотрим **доказательство** леммы 2.5.1. Из **теоремы** 2.5.1.1 имеем

$$\begin{aligned} V_\beta(f^\infty) &= [P^*(f) r(f)] / (1 - \beta) + [I - P(f) + P^*(f)]^{-1} r(f) - \\ &- P^*(f) r(f) + \varepsilon(\beta, f), \end{aligned} \quad (2.41)$$

где $\varepsilon(\beta, f) \rightarrow 0$ при $\beta \rightarrow 1 - 0$.

Определим оператор $L_\beta(f)$ равенством $L_\beta(f)X = r(f) + \beta P(f)X$. В частности, $L_1(f)$ совпадает с оператором $L(f)$.

Лемма 2.5.2. Имеет место следующее равенство

$$L(f)A(f) = u(f) + A(f). \quad (2.42)$$

Докажем это следующим образом:
из определения $A(f)$ и формулы (2.23) имеем

$$\begin{aligned}
 L(f)A(f) &= \lim_{\beta \rightarrow 1-0} [r(f) + \beta P(f)A(f)] = \\
 &= \lim_{\beta \rightarrow 1-0} \left\{ r(f) + \beta P(f) \sum_{i=0}^{\infty} \beta^i [P(f) - P^*(f)]^i r(f) \right\} = \\
 &= \lim_{\beta \rightarrow 1-0} \left\{ r(f) + \sum_{i=0}^{\infty} \beta^{i+1} [P(f) - P^*(f)]^{i+1} r(f) + \beta P^*(f) r(f) \right\} = \\
 &= \lim_{\beta \rightarrow 1-0} \left\{ \sum_{i=0}^{\infty} \beta^i [P(f) - P^*(f)]^i + \beta P^*(f) r(f) \right\} = A(f) + u(f). \tag{2.43}
 \end{aligned}$$

Лемма 2.5.3. Если f^∞ – оптимальная стратегия, то для любого элемента $g \in F$

$$P(g)u(f) < u(f), \tag{2.44}$$

$$P(g)u(f) = u(f) \text{ и } L(g)A(f) \leq L(f)A(f). \tag{2.45}$$

Приведем **доказательство** через следующую теорему.

Теорема 2.5.3.1. Зафиксируем политику $f \in S$.

Тогда: 1. Если $G(f)$ пусто, то $f \in \dot{F}$.

2. Если $G(i, f) \neq \emptyset$ при некотором $i \in S$, то для политики g такой, что $g(i) \in G(i, f)$ и $g(j) = f(j)$ при всех j , для которых $g(i) \notin G(j, f)$, имеет место неравенство $V_\beta(g^\infty) > V_\beta(f^\infty)$ при всех $\beta (< 1)$, достаточно близких к 1.

Доказательство теоремы 2.5.3.1. 1. Если $G(f)$ пусто, то есть если $G(i, f)$ пусто при всех $i \in S$, то $u(f) \geq P(g)u(f)$ или $u(f) - P(g)u(f)$ и $v(f) > r(g) - P(g)u(f) + P(g)v(f) = r(g) - u(f) + P(g)v(f) = r(g) - u(f) +$
 $+ P(g)v(f).$

Таким образом, $V_\beta(g, f^\infty) \leq V_\beta(f^\infty)$ для всех β , достаточно близких к 1, а $V_\beta(f^\infty) \geq V_\beta(g^\infty)$ для всех $g \in F$ и всех β , достаточно близких к 1. В силу (2.29) получаем, что $f \in \dot{F}$.

2. Если $g(i) \in G(i, f)$ для некоторого i и $g(i) = f(i)$, если $g(i) \notin G(i, f)$, то $V_\beta(g, f^\infty) > V_\beta(f^\infty)$ для всех β , достаточно близких к 1.

Эта теорема описывает итерационный алгоритм нахождения стратегий управления, позволяющий отыскать стационарную оптимальную стратегию $f \in \dot{F}$ за конечное число итераций.

Следовательно, из теоремы (2.5.3.1) следует, что если f^∞ – оптимальная стратегия управления, то либо $P(g)u(f) < u(f)$, либо $P(g)u(f) = u(f)$ и

$r(g) + P(g)v(f) \leq u(f) + v(f)$ для любой стратегии g^∞ . В силу (2.42) имеем $L(f)A(f) = u(f) + A(f)$, а из теоремы (2.5.1.1) получаем, что

$$L(f)A(f) = u(f) + v(f) + u(f) \text{ и} \\ L(g)A(f) = r(g) + P(g)A(f) = r(g) + P(g)v(f) + P(g)u(f). \quad (2.46)$$

Таким образом, если $P(g)u(f) = u(f)$, то из неравенства

$$r(g) + P(g)v(f) \leq u(f) + v(f) \text{ следует,} \\ \text{что } r(g) + P(g)v(f) + u(f) \leq u(f) + v(f) + u(f), \\ \text{то есть } L(g)A(f) \leq L(f)A(f).$$

2.6. Схемы принятия оптимальных управленческих решений

Рассмотрим три метода, позволяющих находить или аппроксимировать оптимальный доход v^* , а также строить оптимальные политики управления и политики управления, близкие к оптимальным. **Первый** из них основан на хорошо известном методе **последовательных приближений**, **второй** метод состоит в сведении исходной задачи к эквивалентной ей задаче **математического программирования**, **третий** является обобщением одного из итерационных алгоритмов Ховарда.

Первый метод использует одно лишь условие сжатия и не опирается на свойства монотонности. Итак, пусть выполняется условие сжатия. Тогда рассмотрим и докажем следующую теорему.

Теорема 2.6.1. Пусть выполняется условие сжатия. Тогда неравенство

$$\rho(Au, Av) \leq c\rho(u, v) \quad (2.47)$$

справедливо при любых $u, v \in V$.

Доказательство. Рассмотрим произвольные u, v, x и обозначим $k = (Au)(x) - (Av)(x)$. Разберем сначала случай $k > 0$. Для каждого целого положительного n обозначим d_x^n элемент множества Dx , такой, что $h(x, d_x^n, u) \geq (Au)(x) - k/n$.

$$\text{Ясно, что } (Au)(x) - k/n \geq (Av)(x) \geq h(x, d_x^n, v).$$

Последнее неравенство следует из определения отображения A . Комбинируя оба неравенства, получаем

$$0 \leq (Au)(x) - (Av)(x) - k/n \leq h(x, d_x^n, v) - c\rho(u, v). \quad (2.48)$$

Так как неравенство (2.48) верно при любом n , то $|(Au)(x) - (Av)(x)| \leq c\rho(u, v)$. Случай $k < 0$ рассматривается аналогично. При $k = 0$ неравенство тривиально.

У сжатого отображения A существует единственная неподвижная точка, то есть единственный элемент $v^* \in V$, такой, что

$$v^*(x) = \sup_{d_x \in D_x} h(x, d_x, v^*) \quad (2.49)$$

при любом $x \in \Omega$.

Уравнение (2.49) представляет собой записанное в довольно общем виде уравнение **оптимальности**. Тогда по теореме 2.6.1 $\rho(A^n, v, v^*) \rightarrow 0$ при любом элементе v ; это означает, что v^* может быть найден приближенно путем последовательного применения оператора A , начиная с любого элемента v . На основании полученных выше результатов можно строить оценки для оптимальных политик, но здесь эти вопросы не рассматриваются.

Второй метод состоит в сведении исходной задачи к некоторой задаче математического программирования. Причем в этом случае существенно используется как условие монотонности, так и N -сжатия. Пусть $\min v$ обозначает функцию, значение которой в каждой точке x является минимальным среди значений $v(x)$ всех функций v , удовлетворяющих заданному ограничению. Рассмотрим следующую задачу **математического программирования** [11,12,13].

$$\min v \quad (2.50)$$

$$\text{при ограничении } Av \leq v \quad (2.51)$$

Так как $Af = f$, то f является планом данной задачи. А в силу того, что если $Av \leq v$, то $v \geq f$; если $Av \geq v$, то $v \leq f$, где f – оптимальный план этой задачи. Ясно, что если множество Ω конечно, то (2.50) эквивалентно минимизации $\sum_{x \in \Omega} v(x)$ или любой другой положительной комбинации величины $v(x)$.

Сформулированная выше задача может быть записана в следующей эквивалентной форме: $\min v$ при всех x и d_x .

Третий метод является обобщением ховардовских итерационных алгоритмов нахождения стратегий. Пусть выполнены условия монотонности и N -сжатия, и пусть в определении оператора A из формулы

$$(Av)(x) = \sup_{d_x \in D_x} h(x, d_x, v) \quad (2.52)$$

точная верхняя грань достигается при всех v , то есть $Av = H_\gamma v$ для некоторой политики γ , зависящей от v . Пусть n – произвольное функциональное натуральное число.

Процедура улучшения решения:

1. Взять произвольную начальную политику δ .
2. Вычислить v_δ .
3. Вычислить $u = A^{n-1}v_\delta$ и положить $v = Au = H_\delta u$.
4. Если $\rho(u, v) > \varepsilon$, то заменить δ на γ и перейти к п.2.

Если $\rho(u, v) \leq \varepsilon$, то вычислить v_δ и остановиться. Неравенство $v_\delta \leq A^{n-1}v_\delta = u \leq Au = H_\gamma u \leq v_\gamma$ следует из следующих лемм.

Лемма 2.6.2. Пусть выполняется условие монотонности. Тогда, если $u \geq v$, то $Au \geq Av$. Если $Av \geq v$, то $\{A^n v\} \uparrow$. Если $H_\delta v \geq v$, то $\{H_\delta^n v\} \uparrow$.

Доказательство. В силу (2.52) $Au \geq H_\delta u$. Пусть $u \geq v$. Тогда из неравенства $H_\delta u \geq H_\delta v$ следует неравенство $Au \geq H_\delta v$ при любой политике δ , следовательно, $Au \geq Av$. Если $Av \geq v$, то, применяя последовательно первое утверждение доказываемой леммы, получаем $\{A^n v\} \uparrow$.

Рассмотрим условие сжатия N -го порядка (в дальнейшем N -сжатие), являющееся ослаблением условия сжатия.

Условие N -сжатия. Существует натуральное число N и действительное число C ($0 \leq C < 1$) такие, что при любой политике δ модуль оператора H_δ^N меньше или равен C , и при всех δ модуль оператора H_δ меньше или равен 1.

Обозначим v_δ единственную неподвижную точку сжатия отображения H_δ^N . Легко доказать, что v_δ — единственная неподвижная точка отображения H_δ . Так как модуль H_δ не превосходит единицы, то из неравенства треугольника и определения модуля оператора следует, что $\rho(H_\delta^N v, v) \leq N\rho(H_\delta v, v)$. Отсюда в силу теоремы, которая формулируется так:

пусть выполняется условие сжатия. Тогда для любой политики управления $\delta \in \Delta$ и любой функции $v \in V$ имеет место неравенство

$$\rho(v_\delta, v) \leq \rho(H_\delta v, v)/(1 - c), \quad (2.53)$$

получаем $\rho(v_\delta, v) \leq N\rho(H_\delta v, v)/(1 - c)$.

Определим теперь функцию f равенством $f(x) = \sup v_\delta(x)$. Чтобы показать, что $f \in V$, введем отображение E , определенное на V , с помощью равенства

$$(Ev)(x) = \sup_{\delta} (H_\delta^N v)(x). \text{ Так как } H_\delta^N v \leq Ev \leq A^N v \text{ (последнее неравенство}$$

следует из условия монотонности и определений операторов A и E), то множество значений отображений E содержится в V . Тогда из предыдущего утверждения и условия N -сжатия следует, что E — сжатое отображение. Пусть v^* — единственная неподвижная точка отображения E . Так как $H_\delta^N v^* \leq v^*$, то $v_\delta \leq v^*$ при любой политике δ . Следовательно, $f \leq v^*$ и таким образом, $f \in V$, то есть доказали утверждения:

1. v_δ — единственная неподвижная точка отображения H_δ ;

2. $\rho(v_\delta, v) \leq \rho(H_\delta v, v)N/(1 - c)$;

3. E — сжатое отображение, модуль которого меньше или равен c , следующей теоремы

Теорема 2.6.2.1. Пусть выполняются условия монотонности и N -сжатия.

Тогда: *п.п.* 1,2,3 (см. выше)

4. f – единственная неподвижная точка отображений E и A ;

5. если $v \leq f$, то $\rho(A^N v, f) \leq c\rho(v, f)$.

Доказательство теоремы 2.6.2.1 по *п.п.* 4,5 опирается на следующую лемму

Лемма 2.6.3. Пусть выполняются условия монотонности и N -сжатия. Тогда

если $Av \leq v$, то $v \geq f$; если $Av \geq v$, то $v \leq f$;

$Av_\delta \geq v_\delta$ для любой политики $\delta \in \Delta$;

если $H_\delta v \geq v$, то $v_\delta \geq H_\delta v$.

Докажем эту лемму в такой последовательности, причем докажем только *п.п.* 1, поскольку *п.п.* 2 и 3 очевидны. Пусть 1 – функция, определенная на Ω и принимающая единственное значение, равное 1, то есть $1(x) = 1$ при всех $x \in \Omega$. Предположим сначала, что $Av \leq v$. Тогда $H_\delta v \leq v$ при любой политике δ , откуда следует, что $H_\delta^n v \leq v$ при каждом n . Поэтому $v_\delta \leq v$ для любой политики δ . Следовательно, $f \leq v$.

Далее, так как $Af \geq H_\delta f \geq H_\delta v_\delta = v_\delta$ при любой политике управления δ , то $Af \geq f$. Предположим теперь, что $Av \geq v$. Определим функцию u равенством $u(x) = \max\{v(x), f(x)\}$. Тогда $Au \geq Av$ и $Au \geq Af$, откуда следует, что $Au \geq u$. Для произвольного положительного ε выберем δ , такое, что $H_\delta u \geq Au - \varepsilon 1$. Докажем, что $H_\delta^n(Au) \geq Au - n\varepsilon 1$. Проведем доказательство **индукцией** по n . При $n=1$ утверждение верно, так как $Au \geq u$. Предположим, что оно справедливо при некотором n . Тогда

$$H_\delta^{n+1}(Au) \geq H_\delta(Au - n\varepsilon 1) \geq H_\delta(Au) - n\varepsilon 1 \geq Au - (n+1)\varepsilon 1,$$

то есть утверждение верно и для $n+1$, что завершает доказательство. Так как $v_\delta \leq f \leq u \leq Au$, то из условия N -сжатия вытекает, что $H_\delta^N(Au) \leq v_\delta + c\rho(Au, v_\delta) \cdot 1$. Отсюда получаем $0 \leq Au - v_\delta \leq [N_\varepsilon + c\rho(Au, v_\delta)] \cdot 1$. Предположим теперь, что $Au > f$. Выберем число $\varepsilon = \rho(Au, v_\delta)(1-c)/(2N)$. Тогда $\rho(Au, v_\delta) \leq \rho(Au, v_\delta)(1+c)/2 < \rho(Au, v_\delta)$. Таким образом, мы пришли к противоречию. Следовательно, $Au = f$, а $v \leq f$.

Доказательство теоремы 2.6.2.1 для пунктов 1,2,3 произведено, то есть $Af \geq f$. Тогда $A(Af) \geq Af$ и в силу леммы 2.6.3 получаем $Af \leq f$. Следовательно $Af = f$. Если $Ag = g$, то из леммы 2.6.3 вытекает, что $g = f$. Таким образом, f – единственная неподвижная точка отображения A .

Ранее было установлено, что $f \leq v^*$. Для доказательства равенства $v^* = f$ определим функцию u следующим образом

$$u(x) = \max_{0 \leq n \leq N} (A^n v^*)(x).$$

Тогда $Au \geq A^n v^*$ для $1 \leq n \leq N$. Так как $A^N v^* \geq Ev^* = v^*$, то $Au \geq u$. Следовательно, в силу леммы 2.6.3 выполняется неравенство $u \leq f$, откуда вытекает, что $v^* \leq f$. Значит, $f = v^*$. Для доказательства п.5 теоремы 2.6.2.1 заметим, что $A^N v \geq Ev$. Если $v \leq f$, то $Ev \leq A^N v \leq f = Ef$, откуда следует цепочка неравенств $\rho(A^N v, f) \leq \rho(Ev, Ef) \leq c\rho(u, v)$.

Таким образом, из лемм 2.6.2 и 2.6.3 следует, что $v_\delta \leq A^{n-1} v_\delta = u \leq Au = H_\gamma u \leq v_\gamma$.

2.7. Анализ оптимальных решений с использованием теории игр (стохастические игры)

Стохастическая игра представляет собой случайный процесс с дискретным временем и дискретным пространством состояний, у которого случайные переходы из одного состояния в другое зависят от поведения обоих игроков. Число состояний предполагается конечным. На каждом шаге игра может находиться в одном из состояний N возможных состояний. Игроков, участвующих в игре, будем называть игрок I и игрок II. Если игра находится в состоянии i и игрок I выбирает чистую стратегию k ($1 \leq k \leq m_i$), а игрок II – чистую стратегию r ($1 \leq r \leq n_i$), то с вероятностью P_{ij}^{kr} следующим состоянием игры будет состояние j , и игрок I получил выигрыш a_i^{kr} . Игрок I стремится выбрать стратегию, максимизирующую его суммарный средний выигрыш, накопленный к моменту остановки игры, в то время как игрок II выбирает стратегию, минимизирующую этот выигрыш. Фиксируя i -е состояние игры, приходим к классической прямоугольной игре Γ_i . Таким образом, «стохастическая игра» состоит из набора прямоугольных игр $\Gamma = \{\Gamma_i (i = 1, 2, \dots, N)\}$.

Для стохастических игр введено понятие **остановки**, заключающееся в том, что для каждого состояния i существует положительная вероятность

$$1 - \sum_{j=1}^N P_{ij}^{kr} = S_i^{kr} > 0 \quad (2.54)$$

остановки игры в состоянии i . Соответствующая игра называется **стохастической игрой с остановкой**. Следует отметить, что существуют стохастические игры без остановки, в которых игра никогда не кончается. Для таких игр в качестве критерия можно выбрать средний выигрыш игрока I, приходящийся на одну пару.

В случае, когда один из игроков не имеет возможности выбора стратегии (такой игрок называется **пассивным**), стохастическая игра сводится к марковскому процессу принятия решений. Таким образом, стохастические игры

представляют особый интерес с точки зрения марковских процессов принятия решений.

Стохастическая игра с остановкой задается начальным состоянием и набором из $N^2 + N$ матриц

$$P_y = (P_y^{kr} / k = 1, 2, \dots, m_i, r = 1, 2, \dots, n_i), \quad (2.55)$$

$$A_i = (a_i^{kr} / k = 1, 2, \dots, m_i, r = 1, 2, \dots, n_i), \quad (2.56)$$

элементы которых удовлетворяют условиям

$$\begin{aligned} P_y^{kr} &\geq 0, |a_i^{kr}| \leq M, \\ \sum_{j=1}^N P_y^{kr} &= 1 - S_i^{kr} \leq 1 - S < 1, \\ i, j &= 1, 2, \dots, N. \end{aligned} \quad (2.57)$$

Так как произвольная стратегия в общем случае зависит от всей предыстории игры, то весьма трудно придумать для нее удобное обозначение. Ограничимся стратегиями, не зависящими от предыстории и времени, то есть **стационарными рандомизированными (смешанными) марковскими стратегиями**, задаваемыми наборами из N вероятностных распределений. Например, для игрока I такой набор имеет вид

$$x = (x_1, x_2, \dots, x_N), \quad (2.58)$$

где при каждом

$$x_i = (x_i^1, x_i^2, \dots, x_i^{m_i}) \quad (2.59)$$

является распределением вероятностей. Аналогично можно записать стационарную стратегию у игрока II. Если x_i^k равно 1 при некотором k и 0 при остальных k , то x_i — чистая стратегия.

Рассмотрим сначала матричную игру. Для данной матричной игры B обозначим $Val[B]$ верхнюю цену этой игры, то есть минимальное значение выигрыша игрока I, а $X[B]$ и $Y[B]$ обозначим множества оптимальных стратегий игроков I и II соответственно. Если B и C — две матрицы одинаковой размерности, то легко показать, что

$$|Val[B] - Val[C]| \leq \max_{k,r} |b^{kr} - a^{kr}|. \quad (2.60)$$

Возвращаясь к стохастической игре Γ , определим матрицу $A_i(a)$, (k, r) -элемент которой равен

$$a_i^{kr} + \sum_{j=1}^N P_y^{kr} \alpha_j, \quad (2.61)$$

где $1 \leq k \leq m_i$, $1 \leq r \leq n_i$, а α — $(N \times 1)$ -мерный вектор, i -й элемент которого равен α_i . Выберем произвольный вектор $\alpha^{(0)}$. Определим последовательность векторов $\alpha^{(t)}$

$$\alpha_i^{(t)} = Val[A_i(\alpha^{(t-1)})], \quad (2.62)$$

где $t = 1, 2, \dots$. Покажем, что у $\alpha^{(t)}$ при $t \rightarrow \infty$ существует предел, не зависящий от $\alpha^{(0)}$, причем i -й элемент предельного вектора равен цене игры Γ_i .

Определим преобразование T по формуле

$$\begin{aligned} T\alpha &= \beta, \\ \text{где } \beta_i &= \text{Val}[A_i(\alpha)]. \end{aligned} \quad (2.63)$$

В качестве нормы вектора α выберем

$$\|\alpha\| = \max_i |\alpha_i|. \quad (2.64)$$

Тогда из (2.60) получим

$$\begin{aligned} \|T\alpha - T\beta\| &= \max_i |\text{Val}[A_i(\beta)] - \text{Val}[A_i(\alpha)]| \leq \\ &\leq \max_{i,k,r} \left| \sum_{j=1}^N P_{ij}^{kr} \beta_j - \sum_{j=1}^N P_{ij}^{kr} \alpha_j \right| \leq \max_{i,k,r} \sum_{j=1}^N P_{ij}^{kr} \max_j |\beta_j - \alpha_j| = \\ &= (1 - S) \|\beta - \alpha\|. \end{aligned} \quad (2.65)$$

В частности, $\|T^2\alpha - T\alpha\| \leq (1 - S) \|T\alpha - \alpha\|$. Следовательно, последовательность $\alpha^{(0)}, T\alpha^{(0)}, T^2\alpha^{(0)}, \dots$ сходится, причем предельный вектор φ обладает тем свойством, что $\varphi = T\varphi$, то есть φ – неподвижная точка преобразования T .

Пусть некоторый другой вектор ψ удовлетворяет уравнению $\psi = T\psi$, тогда из (2.65) получаем

$$\|\psi - \varphi\| = \|T\psi - T\varphi\| \leq (1 - S) \|\psi - \varphi\|. \quad (2.66)$$

Следовательно, $\|\psi - \varphi\| = 0$. Значит, φ – единственная неподвижная точка отображения T , причем φ не зависит от $\alpha^{(0)}$.

Чтобы показать, что φ_i – цены игры Γ_i , заметим, что, если игрок I в течение первых t шагов будет придерживаться в игре $\Gamma_i^{(t)}$ оптимальной стратегии, а затем будет играть произвольно, то его выигрыш будет отличаться от цены игры $\Gamma_i^{(t)}$ не более чем на $\varepsilon_t = (1 - S)^t M/S$; то же самое верно и для игрока II. Так как $\varepsilon_t \rightarrow 0$, а цена игры $\Gamma_i^{(t)}$ стремится к φ_i , то отсюда следует, что φ_i является ценой игры Γ_i . Таким образом, доказана следующая теорема.

Теорема 2.7.1. Цена стохастической игры Γ является единственным решением φ системы уравнений

$$\varphi_i = \text{Val}[A_i(\varphi)], \quad i = 1, 2, \dots, N. \quad (2.67)$$

Теорема 2.7.2. Оптимальными стационарными стратегиями игроков I и II в стохастической игре Γ являются соответственно стратегии x^* и y^* такие, у которых $x_i^* \in X[A_i(\varphi)]$ и $y^* \in Y[A_i(\varphi)]$ суть оптимальные стратегии игроков I и II в каждой игре $\Gamma_i (i = 1, 2, \dots, N)$, принадлежащей Γ .

Доказательство. Пусть игра Γ_i заканчивается на t -м ходу, и игрок I вместо φ_i получает выигрыш $a_i^{kr} + \sum P_{ij}^{kr} \varphi_j$. Ясно, что в этом случае применение игроком I стационарной стратегии x^* гарантирует ему выигрыш φ_i .

В исходной же игре Γ_i средний выигрыш за t ходов игрока I, использующего стратегию x^* , будет не меньше, чем

$$\varphi_i - (1-S)^{t-1} \max_{i,k,r} \sum_{j=1}^N P_{ij}^{kr} \varphi_j, \quad (2.68)$$

и, следовательно, этот выигрыш не меньше, чем

$$\varphi_i - (1-S)^t \max_j \varphi_j. \quad (2.69)$$

Поэтому суммарный выигрыш игрока I больше или равен

$$\varphi_i - (1-S)^t \max_j \varphi_j - (1-S)^t M/S. \quad (2.70)$$

Так как все сказанное справедливо при любых достаточно больших t , то отсюда следует, что x^* – оптимальная стратегия для игрока I в игре Γ_i . Аналогично, стратегия y^* оптимальна для игрока II.

Нелинейность оператора Val часто затрудняет получение точных решений с помощью теорем 2.7.1 и 2.7.2. Поэтому желательно выразить выигрыш через стационарные стратегии. Пусть $\overline{\Gamma} = \{\overline{\Gamma}_i\}$ – совокупность игр, чистые стратегии которых являются стационарными стратегиями игры Γ . Их платежные функции $\mathcal{G}_i(x,y)$ должны удовлетворять системе уравнений

$$\mathcal{G}_i(x,y) = x_i A_i y_i + \sum_j x_i P_{ij} y_j \mathcal{G}_j(x,y), \quad i=1,2,\dots,N. \quad (2.71)$$

Эта система имеет единственное решение. В самом деле, для линейного преобразования $Tx y$

$$Tx y \alpha = \beta, \quad \text{где } \beta_i = x_i A_i y_i + \sum_j x_i P_{ij} y_j \alpha_j \quad (2.72)$$

справедливо соотношение

$$\|Tx y \beta - Tx y \alpha\| = \max_i \left| \sum_j x_i P_{ij} y_j (\beta_j - \alpha_j) \right| \leq (1-S) \|\beta - \alpha\|, \quad (2.73)$$

что соответствует неравенству (2.65).

Теорема 2.7.3. Каждая игра $\overline{\Gamma}_i$ имеет седловую точку

$$\min_j \max_x H_i(x,y) = \max_x \min_y H_i(x,y) \quad (2.74)$$

$$i=1,2,\dots,N.$$

Любая стационарная стратегия, оптимальная для всех $\Gamma_i \in \Gamma$, является чистой оптимальной стратегией для всех $\overline{\Gamma}_i \in \overline{\Gamma}$, и наоборот. Векторы цен для Γ и $\overline{\Gamma}$ совпадают.

Доказательство довольно просто следует из теоремы 2.7.2. Отметим, что стратегия x , оптимальная для одной игры Γ_i (или $\overline{\Gamma}_i$), может быть неоптимальной для других игр, принадлежащих Γ (или $\overline{\Gamma}$).

Можно показать, что множество всех оптимальных стационарных стратегий игры Γ является замкнутым выпуклым многогранником. Стохастическая игра с рациональными коэффициентами не обязательно обладает рациональной ценой.

Применяя правило Крамера, находим

$$H_i(x, y) = \frac{\begin{vmatrix} x_1 P_{11} - 1 & x_1 P_{12} y_1 \dots - x_1 A_1 y_1 \dots x_1 P_{1N} y_1 \\ x_2 P_{21} y_2 & x_2 P_{22} y_2 - 1 \dots - x_2 A_2 y_2 \dots x_2 P_{2N} y_2 \\ \dots & \dots \\ x_N P_{N1} y_N & x_N P_{N2} y_N \dots - x_N A_N y_N \dots x_N P_{NN} y_N - 1 \end{vmatrix}}{\begin{vmatrix} x_1 P_{11} y_1 - 1 & x_1 P_{12} y_1 \dots x_1 P_{1i} y_1 \dots x_1 P_{1N} y_1 \\ x_2 P_{21} y_2 & x_2 P_{22} y_2 - 1 \dots x_2 P_{2i} y_2 \dots x_2 P_{2N} y_2 \\ \dots & \dots \\ x_N P_{N1} y_2 & x_N P_{N2} y_N \dots x_N P_{Ni} y_N \dots x_N P_{NN} y_N - 1 \end{vmatrix}}. \quad (2.75)$$

Таким образом, в отличие от теоремы о минимаксе для билинейных форм, равенство (2.74) может не выполняться над произвольным упорядоченным полем.

3. СРАВНИТЕЛЬНЫЙ АНАЛИЗ ОПТИМАЛЬНОГО УПРАВЛЕНИЯ В ПРОЕКТНО-ТЕХНОЛОГИЧЕСКИХ РЕШЕНИЯХ АВИАЦИОННОГО ПРОИЗВОДСТВА

В основе деятельности инженера-проектировщика лежит **процесс проектирования**, под которым в общем случае понимают выбор некоторого способа действия, направленного на составление описания, необходимого для создания в определенных условиях еще несуществующего объекта с возможной оптимизацией заданных его характеристик. Внедрение вычислительной техники в инженерно-технологическую деятельность потребовало строго формального подхода к процессу проектирования сложных технических объектов, к которым относятся технологические процессы в авиаракетостроении и в машиностроении.

При создании систем автоматизированного проектирования (САПР) в основу общего подхода к процессу проектирования закладывается алгоритм, включающий три этапа: **синтез, анализ и принятие оптимального решения** (рис.3.1).

Рассмотрим данный алгоритм (рис.3.1). После определения цели проектирования происходит формирование (генерирование) возможных вариантов (альтернатив) решения проектной задачи. Этот этап называется синтезом и охватывает наиболее творческие виды работ по созданию объекта. В современных САПР на этом этапе могут генерироваться принципиально новые технические решения.

Следующий этап процесса проектирования - это инженерно-технологический анализ. Он направлен на детализацию намеченных вариантов решения задачи (определения структуры и отдельных параметров проектируемого объекта - изделия) и проверку возможных условий функционирования объекта. Определяющими видами работ на данном этапе являются **математическое моделирование** объекта — изделия и его **исследование** на основе этой модели с целью выяснения основных функциональных свойств в рассматриваемой области.

Информация, полученная в результате анализа, позволяет перейти к третьему этапу процесса проектирования - **принятию оптимального решения**. Это наиболее ответственный этап, цель которого — выявление **единственного решения** задачи среди возможных вариантов. На этом этапе наиболее универсальными являются многошаговые методы принятия решений, при которых каждый последующий шаг сужает область поиска и ограничивает число альтернатив.

Рассмотренная схема (рис.3.1) показывает, что основные задачи, решаемые на втором этапе - этапе анализа - связаны с оптимизацией технических объектов (технических решений).

Итак, **оптимизация** - это процесс нахождения **экстремума** некоторой количественной величины (параметра) проектируемого объекта, представляемой в виде функции (функционала). Если эта функция характеризует положительное свойство объекта, то ищется максимальное ее значение, если отрицательное - то минимальное.

Обычно в инженерно-технологической практике используется термин «**оптимальное решение**», или «**оптимальный проект**», под которым в этом случае понимается наилучшее из некоторого множества решение, удовлетворяющее всем требованиям, предъявляемым проектируемому объекту.

Широкое использование во всех сферах инженерной деятельности различных методов и приемов оптимизации, в основе которых лежит определенный математический аппарат, позволило сформулировать целое направление прикладной математики, получившее название «исследование операций».

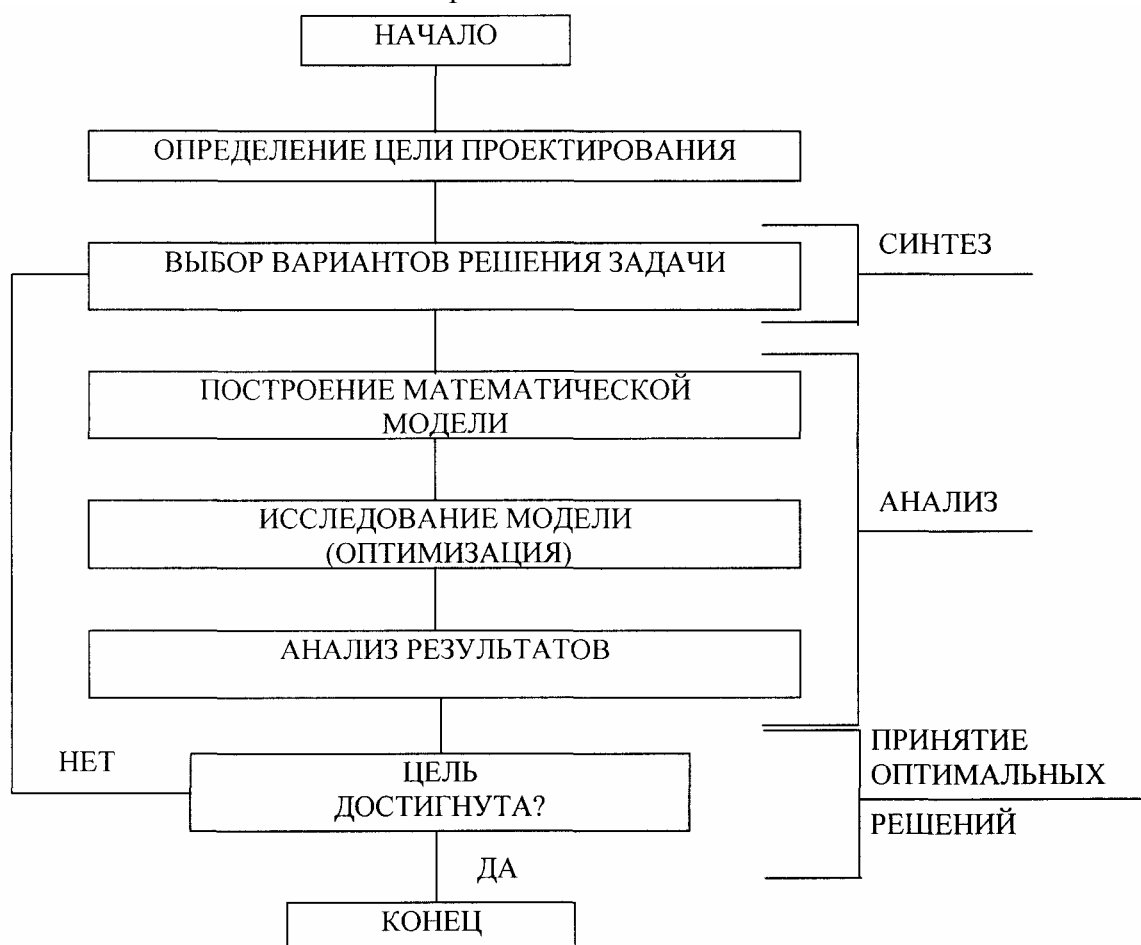


Рис. 3.1. Схема алгоритма процесса проектирования сложного технического объекта

3.1. Теория оптимизации в проектных решениях подготовки авиационного производства

В современном представлении **теория оптимизации** в инженерной практике включает совокупность фундаментальных математических результатов и численных методов, ориентированных на нахождение наилучших вариантов из множества альтернатив и позволяющих избежать полного перебора и сравнения возможных вариантов. Процесс оптимизации лежит в основе инженерной деятельности, направленной на проектирование новых, более эффективных и менее дорогостоящих технических объектов. Достижение этих двух основных целей любого процесса проектирования сопряжено, как уже отмечалось, с синтезом различных элементов, анализом множества их состояний и выбором из них такого состояния, при котором обеспечиваются наилучшие показатели функционирования технического объекта.

Размерность большинства инженерных задач достаточно велика, и проведение расчетов по оптимизации требует значительных затрат времени. Поэтому в условиях неавтоматизированного проектирования, и в частности в технологии авиаракетостроения, решение задач оптимизации практически не проводилось. Становление теории оптимизации во многом связано с появлением сходящихся численных методов оптимизации.

Большинство используемых методов оптимизации являются по своей сути инвариантными и могут использоваться при решении различных проектных задач. Поэтому в настоящее время разработаны десятки численных методов оптимизации, оформленных в виде стандартных процедур (алгоритмов) и хранящихся в библиотеках прикладных программ вычислительных центров, которые открыты для доступа различным пользователям. В этих условиях перед проектировщиком встает задача правильного выбора метода и соответствующих наборов программ. Так, при расчете оптимальных режимов резания могут использоваться стандартные программы **симплекс-метода** или других методов линейного программирования. Однако при этом достаточно сложной задачей является стыковка принятой математической модели объекта с соответствующими программами оптимизации.

Поиск оптимальных технических решений в технологии авиаракетостроения затруднен в связи с низким уровнем формализации существующих методов проектирования технологических процессов и сложностью построения соответствующих математических моделей. Поэтому главным вопросом оптимизации технологических процессов при создании САПР ТП является разработка математических моделей различных технологических объектов и их информационное обеспечение.

Сфера применения методов оптимизации в технологии машиностроения вообще, достаточно широка: проектирование отдельных структурных элементов технических систем, какими, например, являются режимы резания, проек-

тирование более сложных структур, таких, как технологические маршруты и операции обработки, и, наконец, проектирование цехов и промышленных предприятий в целом и т.п.

3.2. Оптимальные структуры математических моделей конструкторско-технологического проектирования

При проектировании самолетов значительное число задач относится к **оптимизационному** классу, а технологический процесс авиационного производства формально можно представить как упорядоченное множество элементов структурной модели, каждый элемент которой выполняет определенную функцию (работу) и находится в конструктивной, функциональной, информационной связи с другими элементами. При этом можно выделить два основных их типа.

Первый тип представляет собой случаи, в которых цель процесса оптимизации может быть выражена как функция F определенного числа переменных (проектных параметров)

$$F = f(x_i), \quad i = \overline{1, n}. \quad (3.1)$$

В этом случае F называется **целевой функцией**, а ее аргументы **параметрами** или **управлениями**.

Можно предположить, что n параметров принадлежат n -мерному **евклидову** пространству, в котором расстояние ρ между двумя точками $X' = (x'_1, x'_n)$ и $X'' = (x''_1, \dots, x''_n)$ определяется следующим уравнением

$$\rho(X', X'') = \left[\sum_{i=1}^n (x'_i - x''_i)^2 \right]^{1/2}. \quad (3.2)$$

Переменные x_i могут быть интерпретированы как компоненты вектора X в n -мерной поверхности управления. В этом смысле вектор X называют вектором управления. С изменением вектора X меняется целевая функция, образуя поверхность в $(n+1)$ -мерном евклидиане, называемую **целевой поверхностью** или **поверхностью отклика**.

На практике в большинстве случаев на отдельные параметры или (и) их комбинацию априори накладывается ряд ограничений.

Задача оптимизации теперь может быть определена как выбор значений x_i (с учетом ограничений), минимизирующих (или максимизирующих) целевую функцию.

Второй тип связан с процессами, которые описываются системой обыкновенных дифференциальных уравнений

$$\frac{dx_i}{dt} = f_i(x_i, u_j), \quad i = \overline{1, n}, \quad j = \overline{1, r}, \quad (3.3)$$

где x_i — так называемые переменные состояния или фазовые переменные;

u_j — переменные управления или параметры управления, влияющие на ход процесса.

Цель оптимизации обычно выражается как требование найти изменение управления по времени t $u_j(t)$, которое обеспечивает **экстремум функционала**, выражающего некоторую целевую функцию F , связанную с переходом объекта от заданного начального фазового состояния $x_i(t_0)=x_{i0}$ к предписанному конечному фазовому состоянию $x_l(t_l)=x_{il}$

$$F = \int_{t_0}^{t_l} f_0(x_1, \dots, x_n; u_1, \dots, u_r) dt = \text{opt}. \quad (3.4)$$

Большинство задач оптимального **проектирования** самолета и его частей относятся к **первому типу**.

Принципиальной задачей **второго типа** является нахождение оптимальной траектории самолета, обеспечивающей минимальное время набора высоты (или снижения) или минимальный расход топлива.

(На практике на ранних этапах проектирования большинство задач второго типа удастся свести к задачам первого типа, поэтому в дальнейшем будем рассматривать в основном лишь их).

Для постановки оптимизационной задачи необходимо:

1. Определить цель оптимизации (выбрать целевую функцию F);
2. Выбрать параметры, наиболее существенно влияющие на целевую функцию (x_i $i=\overline{1, n}$);
3. Выразить целевую функцию через параметры оптимизации $F=f(x_i)$ в такой математической форме, которая, с одной стороны, отражала бы наиболее близко истинное влияние каждого параметра, а с другой, обеспечивала бы наибольшую простоту предстоящего процесса оптимизации;
4. Математически описать все ограничения на отдельные параметры и (или) их комбинацию.

Такие ограничения могут быть выражены в виде равенств

$$\Psi_j(x_i) = B, \quad j=\overline{1, m} \quad (3.5)$$

либо в виде неравенств, когда указываются допустимые верхние и нижние пределы изменения параметра

$$x_i'' \leq x_i \leq x_i^g. \quad (3.6)$$

Кроме этого ограничения типа неравенств могут быть также наложены на комбинацию (функцию) параметров $E_k(x_i)$

$$C_k'' \leq E_k(x_i) \leq C_k^g, \quad k=\overline{1, p}. \quad (3.7)$$

Перечисленные выше ограничения определяют допустимую область изменения проектных параметров X . Эта область в сокращенной записи определяется следующим образом

$$X_{don} = \{X/g_i(X) \leq 0, \quad i=\overline{1, m}\}, \quad (3.8)$$

где $g_i(X)$ – ограничения.

Любой вектор X , принадлежащий допустимой области ($X \in X_{don.}$), определяет допустимый вариант проекта, а совокупность таких векторов образует допустимую область изменения проектных параметров.

Предположим, что в допустимой области изменения проектных параметров (области существования проекта) может существовать проект, параметры которого определяют экстремум целевой функции. Предположим, что ищем минимум целевой функции. Тогда задача оптимального проектирования формулируется следующим образом: найти вектор проектных параметров $X^*=(x_1, x_2, \dots, x_n)$, соответствующий минимуму величины критерия оптимальности при выполнении системы неравенств (3.8). Краткая запись этой задачи $\min F(X)$

$$\begin{aligned} X &\in X_{don.}; \\ X_{don.} &= \{X/g_j(X) \leq 0, j=\overline{1, m}\}. \end{aligned} \quad (3.9)$$

Эта запись означает, что подмножество проектных параметров $X_{don.}$ n -мерного евклидова пространства E^n состоит из всех тех векторов $X=(x_1, x_2, \dots, x_n)$, для которых выполняется заданные ограничения. Запись в виде (4.9) соответствует наиболее сложной задаче нахождения условного экстремума многопараметрической функции. Эта задача является типичной при оптимизации облика самолета и его конструктивно-силовой схемы.

Частным случаем указанной задачи является задача нахождения экстремума действительной функции n действительных переменных, не содержащих ограничений, то есть задача вида

$$\begin{aligned} \min F(X); \\ X \in E^n. \end{aligned} \quad (3.9')$$

Простейшей из задач оптимизации является задача отыскания экстремума действительной функции действительной переменной в заданных границах ее изменения, то есть задача вида

$$\begin{aligned} \min F(x); \\ a < x < b, \end{aligned} \quad (3.9'')$$

где a и b – заданные границы изменения переменной.

В зависимости от постановки оптимизационной задачи, вида целевой функции, наличия и вида ограничений применяют те или иные методы оптимизации.

3.3. Методы оптимизации проектных решений

Методы оптимизации можно разделить на аналитические и численные.

Аналитические методы, базирующиеся на классическом математическом анализе, включают в себя дифференциальное и вариационное исчисления, метод множителей Лагранжа. В зависимости от **вида функции цели и ограничений** подразделяются на классический метод дифференцирования, линейное, квадратичное, выпуклое и динамическое программирование.

Аналитические методы находят применение при решении классических задач и задач с ограничениями в виде уравнений.

Для решения задач без ограничений используют методы исследования производной функции. Путем приравнивания производной нулю отыскиваются точки экстремума, а затем исследуются точки с помощью второй производной для отыскания максимума.

Рекурсивные методы относятся к методам, позволяющим определить одну переменную за одну расчетную операцию. Решение всей задачи осуществляется путем поочередного определения переменных. Наиболее распространенным среди этих методов является динамическое программирование.

Итерационные методы объединяют наибольшую группу методов поиска оптимумов. К ним относятся способы расчета функции цели в одной или нескольких вероятностных точках для определения «лучшей» точки. Расчет выполняют до тех пор, пока не приблизятся к назначенному критерию на расстояние, меньшее некоторого заданного значения. Эти методы позволяют устанавливать только локальные оптимумы, однако они могут применяться в случаях, когда оптимизацию проводят в различных исходных точках.

Стохастические методы оптимизации (методы случайного поиска решений) включают процедуры накопления и обработки информации, в которые сознательно вводится элемент случайности. Преимущества этих методов заключаются в их простоте, надежности, достаточной точности и легкости программирования.

Для использования классических методов определения экстремумов функций и функционалов - дифференциального и вариационного исчисления — в общем случае обязательно также отсутствие ограничений на искомые параметры. Множители Лагранжа используются при решении задач вариационного исчисления и некоторых численных методах оптимизации.

В задачах классического анализа на безусловный экстремум функции одной или многих переменных необходимым условием оптимизации является равенство нулю градиента функции $F(X)$

$$\nabla F(X^*) = \left(\frac{\partial F}{\partial x_1}, \dots, \frac{\partial F}{\partial x_n} \right) = 0. \quad (3.10)$$

При ограничениях в виде равенств оптимальное решение X^* будет определяться соотношением

$$\nabla L(X^*, \lambda^*) = 0, \quad (3.11)$$

где $L(X) = F(X) + \lambda^T g(X)$ — функция Лагранжа,

$\lambda^T = (\lambda_1, \dots, \lambda_m)$ — вектор неопределённых множителей.

Достаточные условия оптимальности связаны с понятием матрицы Гессе (гессиана) H , представляющей собой матрицу вторых частных производных функций $F(X)$ [21,22,23]

$$H = \left\| \frac{\partial^2 F(X)}{\partial x_i \partial x_j} \right\|, \quad i, j = \overline{1, n}. \quad (3.12)$$

Чтобы X^* была точкой минимума, достаточно положительной полуопределенности (определенности) гессиана $H(X^*)$, то есть выполнения условия $\Delta X' H(X^*) \Delta X \geq 0$ для любых $\Delta X = 0$.

Понятия градиента и гессиана используются во многих алгоритмах поиска экстремума. Для многоэкстремальной функции $F(X)$ все корни уравнений (3.10) и (3.11) необходимо еще проверить на глобальный оптимум.

Методы дифференциального исчисления применяются при решении задач частной оптимизации при исследовании сравнительно несложных моделей с одним-тремя переменными, границы допустимых значений которых не проходят через экстремум целевой функции.

Если целевая функция представлена в виде функционалов, используют методы вариационного исчисления. Эти методы находят применение при решении задач оптимизации параметров траектории полета и режима изменения тяги двигателя по времени. В этом случае задача оптимизации параметров сводится к интегрированию системы дифференциальных уравнений Эйлера. В качестве других примеров использования методов вариационного исчисления при решении оптимизационных задач проектирования самолетов назовем задачи об оптимальном законе изменения относительной толщины крыла по размаху и об оптимальной крутке серединой поверхности крыла сверхзвукового самолета для получения максимального аэродинамического качества.

Важным свойством задач линейного программирования является выпуклость целевой функции допустимого множества. Функция $F(X)$ выпукла на выпуклом множестве X , если для любых двух точек

$$\begin{aligned} \dot{x} \in X \text{ и } \ddot{x} \in X \text{ справедливо неравенство} \\ F[\dot{x} - \lambda(\ddot{x} - \dot{x})] \leq F(\dot{x}) + \lambda[F(\ddot{x}) - F(\dot{x})] \\ \text{для } 0 \leq \lambda \leq 1. \end{aligned} \quad (3.13)$$

$$\begin{aligned} \text{Множество } X \text{ выпукло, если из условий } \dot{x} \in X \text{ и } \ddot{x} \in X \text{ следует} \\ x = \dot{x} + \lambda(\ddot{x} - \dot{x}) \in X, \quad \text{где } 0 \leq \lambda \leq 1. \end{aligned} \quad (3.14)$$

Как известно, для решения задачи, если она невырождена, при выполнении условий (3.13) и (3.14) достаточно применить необходимые условия оптимальности.

Среди общих методов наибольшее распространение получил **симплекс-метод** и его разновидности [26]. К специальным методам принадлежат **метод линейных ветвлений**, **распределительный метод** и др. В последнее время разработаны **комбинаторные методы** решения задач целочисленного программирования, наиболее эффективный из них – **метод ветвей и границ**.

Рассмотрим задачу общего вида для выяснения сущности метода, где необходимо найти

$$\min F(x) \text{ при } x \in X, \quad (3.15)$$

где X – конечное множество.

В основе метода ветвей и границ лежит разбиение по определенному правилу множества X на подмножества X_i (ветвление) так, чтобы $X = \bigcup X_i$, и определения нижних оценок (границ) $\xi(X_i)$ целевой функции $F(X)$ на этих подмножествах, удовлетворяющих условию $\xi(X_i) < F(x)$ для всех $x \in X_i$. Если при этом окажется, что для некоторого вектора $X^* \in X_v$ будут выполняться условия $F(X^*) = \xi(X_v) \leq \xi(X_i)$ для всех i , то в силу определения оценок X^* – оптимальное решение задачи (3.15).

Согласно принципу **оптимальности** принятия решений по управлению сложной системой, проблему разбивают на ряд последовательных шагов (этапов), на каждом из которых решают свою оптимизационную задачу. При этом задача минимизации функции многих переменных сводится к последовательному (поэтапному) решению задач минимизации функции одной переменной согласно следующей схеме

$$\min_{x_1, x_2, \dots, x_n} F(x_1, x_2, \dots, x_n) = \min_{x_2} \{ \min_{x_n} [\dots \min_{x_1} F(x_1, x_2, \dots, x_n)] \}.$$

Динамическое программирование ориентированно, прежде всего, на решение **оптимизационных задач**, связанных с исследованием систем, характеристики которых зависят от времени (динамических систем).

В общем случае задачи (3.9), в которых и целевая функция, и ограничения представляют собой нелинейные зависимости от X , называемые задачами **нелинейного математического программирования**, не имеют единого алгоритма или класса алгоритмов их решения. Если для выпуклой функции $F(X)$ и выпуклого множества $X_{доп}$ локальный минимум одновременно является и глобальным, то для невыпуклой области $X_{доп}$ даже при выпуклой функции $F(x)$ задача нелинейного программирования может оказаться многоэкстремальной. Так, если $F(x)$ – квадратичная функция, а ограничения являются линейными функциями, то задача (3.9) называется задачей **квадратичного программирования**.

Разработана методика оптимизации для задач нелинейного программирования, в которых целевая функция и ограничения являются полиномами относительно неотрицательных переменных $x_1, x_2, \dots, x_n$ с положительными коэффициентами следующего вида

$$F(X) = \sum_{i=1}^k P_i(X),$$

где $P_i(X) = c_i x_1^{a_{i1}} x_2^{a_{i2}} \dots x_n^{a_{in}}$,

$c_i > 0$, $x_j > 0$, $j = 1, n$, а показатели степени $a_{i1}, a_{i2}, \dots, a_{in}$ – произвольные вещественные числа.

Такие задачи, называемые задачами **геометрического программирования**, часто встречаются в практике технического автоматизированного проектирования.

Рассмотрим методы решения задачи (3.9) и методы решения частных задач вида (3.9'), то есть задач **безусловной оптимизации**.

В случае выпуклости целевой функции можно найти оптимальное решение X^* , выполнив необходимое условие оптимальности: равенство нулю градиента функции, то есть $\nabla F(X^*)=0$. Однако в большинстве практических задач выпуклость либо не имеет места, либо не поддается проверке. В этом случае для решения задачи рекомендуется выбрать несколько начальных точек поиска и затем взять наинизшую из всех конечных точек.

Такая стратегия рекомендуется для любого алгоритма оптимизации с ограничениями и без них, когда условия, гарантирующие минимум, либо не выполняются, либо проверяются с трудом.

Рассмотрим численные методы решения задачи (3.9'), в основе которых лежат процедуры систематического получения последовательности векторов, то есть точек $X^0, X^1, \dots, X^k$ в E^n , таких, что $F(X^0) > F(X^1) > \dots > F(X^k) > \dots$, и оценки их эффективности при локализации точек минимума. Такие методы получили название методов спуска. Схема алгоритма вычислений при таких методах представлена на рис.3.2.

Решение задачи начинают с выбора начальной точки X^0 . При этом стремятся использовать любую имеющуюся информацию о поведении функции $F(X)$ с тем, чтобы выбрать X^0 как можно ближе к точке минимума. После выбора начальной точки необходимо решить, в каком направлении располагать следующие точки и какой величины шаг должен быть сделан в выбранном направлении.

При любом методе спуска последовательность X^k подчиняется условию $X^{k+1} = X^k + t^k d^k$, $k=0, 1, \dots$, где d^k есть направление, а $\|t^k d^k\|$ – величина шага.

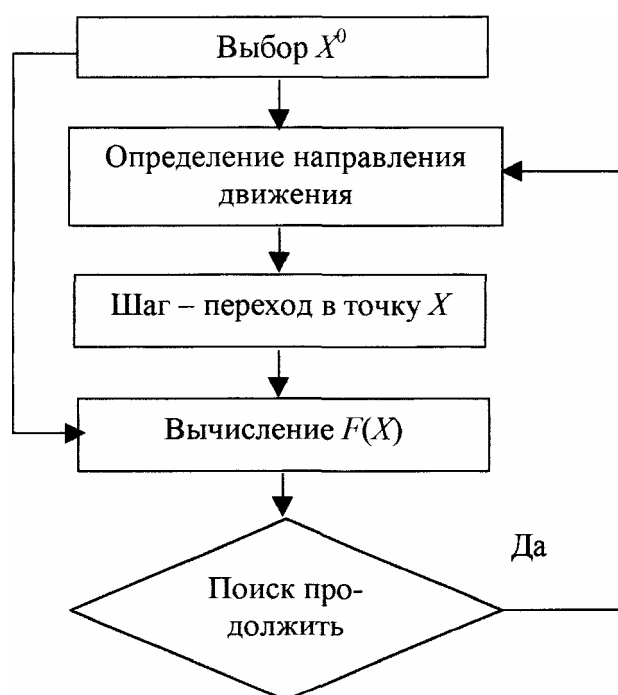


Рис.3.2. Схема алгоритма поиска экстремума

Если d^k нормализовано $\|d^k\|=1$, то величина шага будет равна $|t^k|$.

Механизм образования последовательности точек и его эффективность в локализации точки минимума в сильной степени зависит от вида минимизируемой функции, а также от банка знаний или информации, которая может быть использована для следующего прогноза точки минимума. Изменяя процедуру выбора d^k и t^k , можно варьировать методы спуска.

В зависимости от способа выбора направления очередного шага методы спуска подразделяют на и три группы:

методы, использующие при выборе направления информацию только о значениях функций; их называют методами нулевого порядка или прямыми методами;

методы, которые требуют вычисления вторых производных функций, называют методами второго порядка;

методы, использующие, кроме того, первые производные функций, называют методами первого порядка.

Выбор величины и шага во многом определяет эффективность поиска экстремума, в том числе потребное число шагов (итерации), от которого зависят количество и затраты использования вычислительной техники. Среди многих способов выбора шага наиболее эффективны основные на минимизации функции $F(X)$ в выбранном направлении, то есть решении задачи (3.9") **однопараметрической оптимизации**.

Графическая интерпретация прямых методов и методов, использующих значения производных, представлена на рис.3.3 [23].

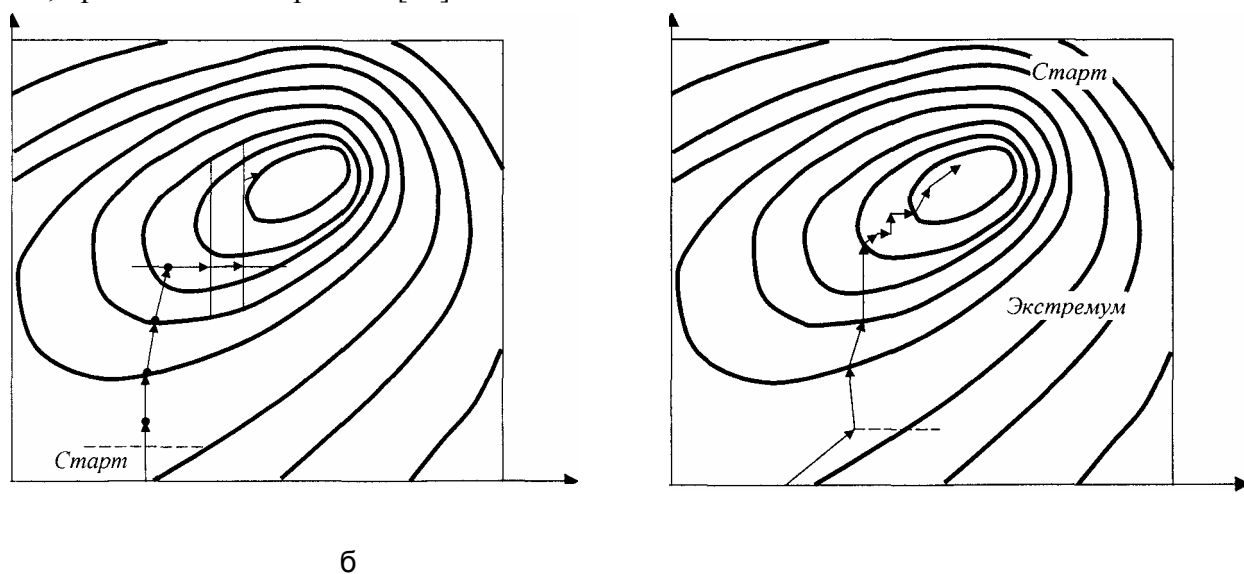


Рис. 3.3. Графическая интерпретация двух методов поиска экстремума: а -прямые методы; б - методы, использующие значения производных;

—> - принятый шаг; - - - исследованный шаг

Видно, что они отличаются выбором направления поиска [21]. В методах первого порядка направление движения на каждом шаге совпадает с опти-

градиентом функции. В этих методах, называемых **градиентными**, точка X^{k+1} выбирается по отношению X^k в направлении $-r^k$, то есть

$$X^{k+1} = X^k - t r^k,$$

где $r^k = \nabla F(X^k)$ – градиент функции $F(X)$ в произвольной точке X^k ; t – некоторая положительная скалярная величина, называемая шагом поиска.

Подбирая величину t , можно обеспечить выполнение условия $F(X^{k+1}) < F(X^k)$. В зависимости от способа выбора шага t существуют разновидности градиентных методов, среди которых наиболее известны **метод градиентного спуска** и **метод наискорейшего спуска**.

В первом методе задают шаг достаточно малой величины и вычисляют градиент на каждом шаге (см. рис.3.3,б). Это может привести к чрезмерно большим затратам (при расчетах на вычислительной технике) объема памяти и машинного времени. В методе наискорейшего спуска величину шага выбирают оптимальной путем решения задачи одномерной минимизации из условия

$$F[X^k - t \nabla F(X^k)] = \min_{t \geq 0} F[X^k - t \nabla F(X^k)].$$

Графическая интерпретация метода представлена на рис.3.4.

К методам первого порядка относится и ряд других методов, в частности весьма эффективный метод сопряженных направлений и его частный случай – метод сопряженных градиентов.

При градиентных методах процесс поиска экстремума прекращается, если для всех компонент вектора градиента $\nabla F(X^k)$ выполняются условия

$$\left| \frac{\partial F(X^k)}{\partial X_j} \right| \leq \delta, \quad j = \overline{1, n},$$

где δ – наперед заданное число, характеризующее точность минимума.

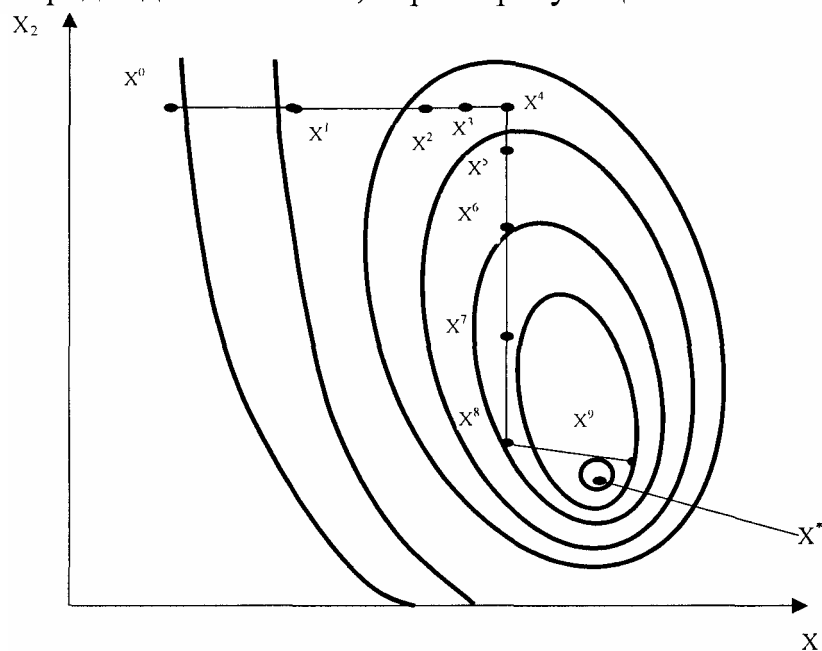


Рис.3.4. Поиск экстремума методом наискорейшего спуска

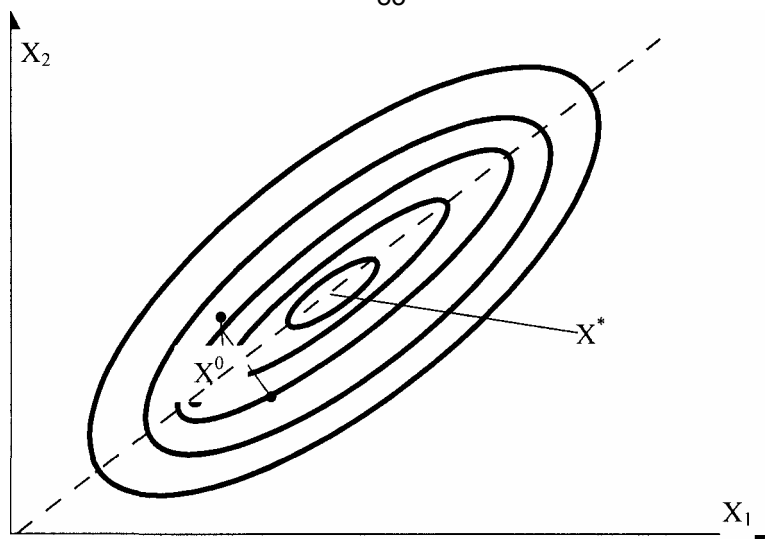


Рис 3.5 Градиентный поиск вдоль «оврага»

Существенным недостатком градиентных методов является их слабая сходимость в случае, если оптимум лежит в «овраге» или на длинном узком «ребре» (рис.3.5). В этом случае траектория градиентных методов характеризуется достаточно медленным зигзагообразным движением вдоль дна «оврага» или «ребра» к точке минимума.

Существует ряд приемов, позволяющих избежать «овражных» эффектов, например, путем изменения масштабов независимых переменных, что изменяет топографию поверхности отклика целевой функции (траектории линий равного уровня). Однако наиболее эффективным является переход к методам второго порядка, поскольку вторые производные минимизируемой функции чувствительны к кривизне линий уровня. Эти методы являются обобщением **метода Ньютона**, который может быть интерпретирован как метод последовательного поиска точек минимума квадратичных аппроксимаций функции $F(X)$. Алгоритм метода Ньютона имеет вид

$$X^{k+1} = X^k - H^{-1}(X^k) \nabla F(X^k),$$

где $H^1(X^k)$ - матрица Гессе целевой функции, вычисленная в точке X^k .

Ясно, что если функция $F(X)$ строго квадратична, метод Ньютона дает решение задачи за один шаг. Однако и в общем случае неквадратичной функции $F(X)$ этот метод потенциально более эффективен при отыскании минимума, чем градиентные методы, поскольку квадратичная функция локально более точно аппроксимирует функцию $F(X)$, чем линейная, лежащая в основе градиентных методов.

Для уменьшения трудностей, связанных с вычислением вторых производных, в методе **переменной методики** (метод Флетчера-Пауэла) вычисление минимума ведется по следующей формуле

$$X^{k+1} = X^k - tH^{-1}(X^0) \nabla F(X^k).$$

При этом матрицу Гессе вычисляют по приближенной формуле без определения вторых частных производных. Величина шага определяется одномерной минимизацией целевой функции на луче

$$H^{-1}\nabla F(X^k).$$

В методе Ньютона - Рафсона для обеспечения сходимости от начального приближения в алгоритм минимизации помимо определения направления поиска вводится процедура выбора длины шага вдоль него

$$X^{k+1} = X^k - t^k H^{-1}(X^k) \nabla F(X^k).$$

Общим для первого и второго методов является то, что все они предусматривают поиск локальных минимумов на последовательности направлений (лучей). Этот поиск является составной частью общей процедуры минимизации. Поиск локального минимума целевой функции $F(X)$ вдоль заданного направления представляет собой задачу однопараметрической минимизации (3.9"). Из разработанных алгоритмов решения этой задачи заслуживают внимание три из них: **метод золотого сечения, метод Фибоначчи и метод хорд.**

Первые два метода не требуют вычисления производной функции. Их основная идея - уменьшить количество вычисления функций и заключить минимум в последовательно убывающие вложенные интервалы $[a, b]$. С этой целью функцию вычисляют в двух внутренних точках I_L (левая) и I_R (правая) текущего интервала $[a, b]$. В качестве нового меньшего интервала $[a_{I+1}, b_{I+1}]$ берется либо $[a, I_L]$, либо $[I_R, b]$. Оба метода основаны на том, что оставшаяся внутренняя точка I_L или I_R интервала $[a_{I+1}, b_{I+1}]$ используется на следующем шаге как одна из внутренних точек этого интервала, с которыми поступают так же, как $[a, b]$. Оба метода применяются для оптимизации **унимодальных** (имеющих один экстремум) функций.

Метод хорд основан на использовании дискретного варианта метода Ньютона для отыскания нуля производной $F'(X)$ функции $F(X)$ в интервале $[a, b]$, если он существует. Собственно метод Ньютона состоит в линеаризации $F(X)$ в окрестности текущей точки, в выборе нуля этой линеаризованной функции в качестве исходной точки следующей итерации.

Существуют методы нулевого порядка, которые конкурируют с методами первого и второго порядков в случае, когда целевая функция $F(X)$ обладает несколькими локальными экстремумами.

Характерной особенностью прямых методов является их эвристический характер, отсутствие строго обоснования. Известные методы этой группы: **метод покоординатного спуска** (метод Гаусса - Зейделя), **метод конфигураций** (метод Хука и Дживса), **метод деформируемого многогранника** (метод Нел-дера и Мида), **метод Розенброка, метод Пауэла** [20,21,22,23].

Наиболее простым из них является **метод покоординатного спуска**, сущность которого состоит в последовательной минимизации целевой функции по отдельным переменным.

Метод конфигурации по своей сути является достаточно общим методом прямого поиска. Согласно этому методу для выбранной начальной точки

X^0 путем изменения одной или нескольких значений компонент вектора X^0 обследуется ее окрестность (см. рис. 3.3,а). После нахождения приемлемого направления вычисляют функцию при постепенно увеличивающемся шаге до тех пор, пока это позволяет определять точки X с меньшими значениями функции $F(X)$ и т.д.

Метод деформируемого многогранника является модификацией метода конфигураций. Здесь функция p переменных $F(X)$ минимизируется с использованием $(n+1)$ вершин некоторого деформируемого многогранника в пространстве этих переменных R^n . Вершина (точка) в R^n , в которой $F(X)$ максимально, проектируется через «центр тяжести» оставшихся вершин. Улучшенные (более низкие) значения целевой функции находятся последовательной заменой точки с максимальным значением $F(X)$ на точки с меньшим значением функции, пока не будет найден минимум. Формально алгоритм записывается следующим образом:

1. Формирование в n -мерном пространстве многогранника

$$a_{i1}^k = [a_{i1}^k, \dots, a_{ij}^k, \dots, a_{in}^k], \quad i = \overline{1, n+1}; \quad k = \overline{0, 1, \dots},$$

где i – вершина на каждом этапе поиска.

2. Определение среди вершин «наихудшей» a_h и «наилучшей» a_l

$$f(a_h^k) = \max\{f(a_1^k), \dots, f(a_{n+1}^k)\};$$

$$f(a_l^k) = \min\{f(a_1^k), \dots, f(a_{n+1}^k)\}.$$

3. Определение «центра тяжести» всех вершин, исключая a_h

$$a_{n+2}^k = \frac{1}{n} \left[\sum_{i=1}^{n+1} a_{ij}^k - a_{hj}^k \right], \quad j = \overline{1, n},$$

где j – координатное направление.

4. Отражение «наихудшей» вершины

$$a_{n+3}^k = a_{n+2}^k + \alpha(a_{n+2}^k - a_h^k), \quad \alpha < 0.$$

5. Растяжение:

$$\begin{aligned} &\text{если } f(a_{n+3}^k) \leq f(a_l^k), \\ &\text{то } a_{n+4}^k = a_{n+2}^k + \gamma(a_{n+3}^k - a_{n+2}^k), \quad \gamma > 1; \\ &\text{если } f(a_{n+4}^k) < f(a_l^k), \text{ то } a_n^k = a_{n+4}^k, \\ &\text{в противном случае} - a_n^k = a_{n+3}^k. \end{aligned}$$

6. Сжатие:

$$\begin{aligned} &\text{если } f(a_{n+3}^k) > f(a_l^k) \text{ для всех } i \neq h, \\ &\text{то } a_{n+5}^k = a_{n+2}^k + \beta(a_n^k - a_{n+2}^k), \quad 0 < \beta < 1, \\ &a_n^k = a_{n+5}^k. \end{aligned}$$

7. Если $f(a_{n+3}^k) \leq f(a_l^k)$ хотя бы для одного $i \neq l$, то вершина a^k заменяется на вершину a_{n+3}^k .

8. Редукция:

если $f(a_{n+3}^k) > f(a_n^k)$, то формируется новый многогранник с уменьшенными вдвое сторонами и вершинами $a_i^k = a_i^k + 0,5(x_i^k - x_i^k)$ и поиск повторяется.

Критерий окончания поиска

$$\left\{ \frac{1}{n+1} \sum_{i=1}^{n+1} [f(a_i^k) - f(a_{n+2}^k)]^2 \right\} \leq \varepsilon.$$

В методе Розенброка, который также является модификацией метода покоординатного спуска, для преодоления сложностей поиска минимума при наличие узких «оврагов» используется процедура поворота ортогональных координатных осей так, чтобы одна из них совпадала с направлением «оврага». Траектория поиска экстремума по этому методу показана на рис.3.6.

В более сложном **методе Пауэла** поиск экстремума производится вдоль сопряженных, а не ортогональных направлений. Также усовершенствована процедура линейного поиска, позволяющая отыскивать не локальный, а глобальный оптимум. Особую группу среди прямых методов занимают **методы случайного поиска**.

Среди методов случайного поиска наиболее эффективен **поиск по случайным точкам** и **поиск по случайным лучам**.

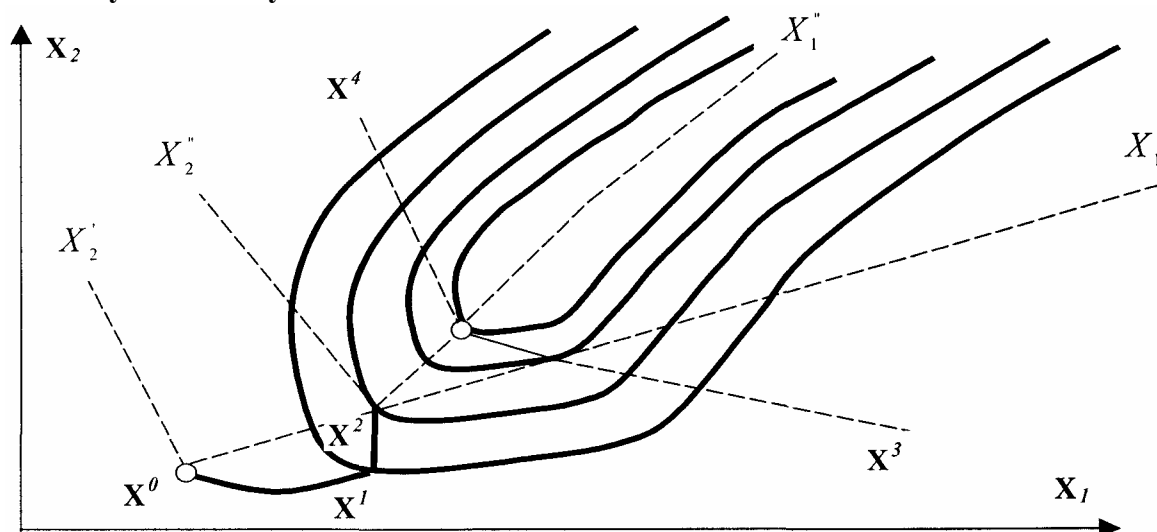


Рис.3.6.Траектория поиска экстремума по методу Розенброка

Простейший алгоритм случайного поиска заключается: в точке X^k формируется случайное направление с помощью единичного случайного вектора ξ^k , равномерно распределенного по всем направлениям. Делается пробный шаг в этом направлении $h_{пр} \xi^k$. Рабочий шаг формируется из условий

$$\Delta X^k = \begin{cases} h_{раб} \xi^k, & \text{если } F(X^k + h_{пр} \xi^k) < F(X^k), \\ 0 & \text{в противном случае.} \end{cases}$$

В результате получаем новое приближение $X^{k+1} = X^k + \Delta X^k$.

Величина $h_{\text{раб}}$ подбирается экспериментально.

Рассмотрим способы решения задачи **нелинейного программирования с ограничениями**, то есть задачи (3.9), представляющей наибольший интерес для проектировщика. Существует два подхода при решении задачи.

Первый основан на непосредственном учете ограничений. Такой подход лежит в основе **метода возможных направлений** (метод Зойтендейка), **метод проективного градиента**, методов **аппроксимирующего линейного программирования**.

Второй подход к решению задачи (3.9) основан на преобразовании задач с ограничениями к более простым задачам без ограничений. Этой цели можно достигнуть **заменой переменных** либо **видоизменением целевой функции** с помощью некоторых функций ограничивающих уравнений. Наиболее распространенным является метод **штрафных функций**, основная цель которого заключается в сведении задачи на условный экстремум к последовательности задач безусловной оптимизации путем использования функций штрафов.

Кратко о некоторых вышеназванных методах условной оптимизации:

1. Метод возможных направлений. Целевую функцию минимизируют как функцию без ограничений до тех пор, пока не встретятся ограничения. Затем находят направление поиска, позволяющее уменьшить целевую функцию, не нарушая ограничений. Такое направление называется допустимым (возможным). Обозначим его S^k , тогда новая точка будет определяться по соотношению

$$X^{k+1} = X^k + h^k S^k.$$

Тогда X^{k+1} , а следовательно и направление S^k будут допустимыми, если выполняется условие

$$\frac{\partial g^T(X^k)}{\partial X} S^k \leq 0. \quad (3.16)$$

Решая задачу

$$\min \nabla F^T(X^k) S^k \quad (3.17)$$

при условии (3.16), определяют вектор допустимых направлений S^k , вдоль которого целевая функция имеет наибольшую скорость убывания. Величину шага h^k вдоль выбранного направления можно определить, решая задачу однопараметрической минимизации вида

$$\begin{aligned} F(X^k + h^k S^k) &= \min F(X^k + h S^k) \\ g(X^k + h S^k) &\leq 0, \quad h \geq 0. \end{aligned} \quad (3.18)$$

2. Метод проектного градиента является модификацией метода возможных направлений. В отличие от п.1 в этом методе при попадании точки X^k в район ограничения допустимое направление поиска S^k определяется не с помощью решения задачи (3.17), а проектированием **антиградиента** — $\nabla F(X^k)$ на

многогранник, являющийся линейной аппроксимацией допустимого множества вблизи точки X^* . Это позволяет учитывать ограничения, как в виде неравенств, так и равенств.

3. Метод аппроксимирующего линейного программирования заключается в сведении задачи нелинейного программирования к задаче линейного программирования путем замены нелинейной целевой функции и функции ограничений последовательностью аппроксимирующих линейных функций. В одних алгоритмах эта цель достигается путем использования линейной интерполяции нелинейных функций, в других — их разложением в ряд Тейлора в окрестности точки X^* .

4. Метод штрафных функций заключается в том, что задача условной оптимизации сводится к эквивалентной задаче безусловной оптимизации путем преобразования целевой функции. Новая целевая функция $\dot{F}(X)$ образуется путем добавления к целевой функции $F(X)$ функции штрафа, составленной из ограничивающих условий таким образом, что приближение к границе допустимой области приводит к резкому увеличению новой целевой функции, то есть нарушение ограничений штрафуются ухудшением $\dot{F}(X)$.

В зависимости от того, находится ли решение задачи на безусловный экстремум внутри или вне исходной допустимой области, различают два типа алгоритмов решения задач методом штрафных функций - алгоритм **внутренней** штрафной функции и алгоритм **внешней** штрафной функции. В первом случае поиск оптимума должен начинаться из допустимой области и его траектория полностью будет лежать внутри этой области. Это достигается при образовании новой целевой функции, например, вида

$$\dot{F}(X) = F(X) + R_k \sum_{i=1}^m 1/g_i(X), \quad (3.19)$$

где $R_k > 0$ - весовой коэффициент.

При приближении к границе допустимой области изнутри какой-либо из элементов вектора ограничений стремится к нулю, а следовательно, функция штрафа приближается к бесконечности. Недостатками этого алгоритма являются необходимость выбора исходной точки внутри области существования, а также его неприменимость при ограничениях в виде равенств.

Во втором случае поиск может начинаться из любой точки, в том числе находящейся вне допустимой области. При этом функция выбирается таким образом, чтобы значения новой целевой функции в допустимой области точно или приближенно равнялись значениям исходной целевой функции, а вне ее - существенно превосходили значениям $F(X)$. Возможный вид такой новой целевой функции

$$\dot{F}(X) = F(X) + R_k \sum_{i=1}^m \delta_i g_i(X), \quad (3.20)$$

$$\text{где } \delta_i = \begin{cases} 0, & \text{если } g_i(x) \leq 0 \\ 1, & \text{если } g_i(x) > 0. \end{cases}$$

Величина штрафа зависит от выбора весового коэффициента R_k . чем он больше, тем ближе $F(X)$ к $F(X)$, тем точнее решение. Однако необходимо иметь в виду, что увеличение R_k ведет к росту роли ошибок счета и, что самое важное, к усложнению поиска экстремума. Это связано с тем, что введение штрафа «искривляет» целевую функцию, образуя двухсторонний «овраг» при ограничениях в виде **равенств** и односторонний «утес» для ограничений в виде **неравенств**. Вследствие этого формулировка ограничений в виде неравенств предпочтительна при решении задачи методом **штрафных функций**.

В силу указанных причин рассмотренный метод обычно применяется для получения приближенных решений при небольших значениях весовых коэффициентов R_k .

Рассмотренные выше методы оптимизации применяют при исследовании детерминированных (неслучайных) функций и процессов, однако в практике проектирования приходится решать оптимизационные задачи, в которых необходимо учитывать случайные факторы. Такие задачи решают методами **стохастического программирования**.

В практике проектирования самолетов могут иметь место задачи оптимизации одновременно по нескольким показателям качества. Например, перед проектировщиком поставлена задача получить наилучшие значения для нескольких характеристик самолета, например, максимизировать дальность полета, минимизировать потребную длину взлетно-посадочной полосы и взлетную массу самолета. Как правило, эти характеристики, выбираемые в качестве критериев, противоречивы и оптимизация по каждому из них привела бы к разным значениям проектных параметров X . В тех случаях, когда не удастся найти обобщенный показатель качества, включающий в себя указанные частные показатели, возникает задача **многокритериальной** (векторной) оптимизации. Для многокритериальной задачи в общем случае решение не является оптимальным ни для одного из частных случаев. В то же время оно является компромиссным для векторного критерия

$$F(X)=[F_1(X),\dots,F_m(X)]. \quad (3.21)$$

Такое решение называется областью компромиссов или областью решений, **оптимальных по Парето**. Для определения минимума по Парето необходимо перейти от задачи векторной оптимизации к задаче нелинейной оптимизации со специально сконструированной скалярной целевой функцией, решив предварительно задачу свертывания векторного критерия оптимальности.

Способы свертывания векторного критерия оптимальности зависят от информации о степени сравниваемости частных критериев оптимальности.

3.4. Предложения по выбору и реализации методов оптимизации при решении проектных задач

Когда поставлена оптимизационная задача перед инженером-проектировщиком встает проблема выбора метода, по возможности однозначно определяющего порядок операций, приводящих к решению. Кратко описанные выше и некоторые другие численные методы оптимизации, оформленные в виде стандартных процедур и хранящиеся в базе данных ЭВМ современных проектно-конструкторских организаций, являются достоянием любого инженера-проектировщика. Необходимо среди имеющихся в информационной базе методов выбрать наиболее приемлемый для решения конкретной проектной задачи и затем состыковать программу с моделью. Разумеется, уже сама постановка проектной задачи на основе инженерного анализа или функционально-стоимостной инженерии позволяет отбросить некоторые методы как неприемлемые. Методы решения различны для задач с любой нелинейностью и с малым числом переменных и задач с резко выраженной нелинейностью с малым и большим числом переменных; задач с одним экстремумом и со многими локальными экстремумами. Концепция применения программ оптимизации по принципу «черного ящика» может привести к результатам решения, весьма далеких от оптимальных.

В развитых системах автоматизированного проектирования, таких как CATIA, CIMATRON, CADD5, UNIGRAPHICS и других имеется специальная система (или подсистема) оптимизации, представляющая собой сложный программный комплекс (набор взаимосвязанных файлов с правилами обмена данными по используемым методам оптимизации и процедурами производства расчетов). Комплекс включает в себя комбинацию различных методов поиска экстремума, объединенных специальной программой - эмулятором, осуществляющей переход к различным методам в зависимости от поведения **целевой функции** на различных шагах итерации [45].

Такой программно-технический комплекс призван обеспечить возможность эффективного использования алгоритмов оптимизации, включенных в систему, путем реализации процессов адаптации поиска, предусматривающих своевременную смену алгоритмов при решении проектной задачи. Он обеспечивает работу как в автоматическом, так и в интерактивном режиме, позволяя пользователю вносить оперативные изменения и в модель, и в исходные данные, а также использовать различные эмпирические приемы, ускоряющие сходимость. Управляющий язык программно-машинного комплекса составляется (подбирается) так, что позволяет использовать **мнемонику** и смысловые конструкции при формировании заданий проектирования и работе с системой пользователей, не владеющих языками программирования. Структура обобщения (укрупнения) такого комплекса представлена на рис.3.7.

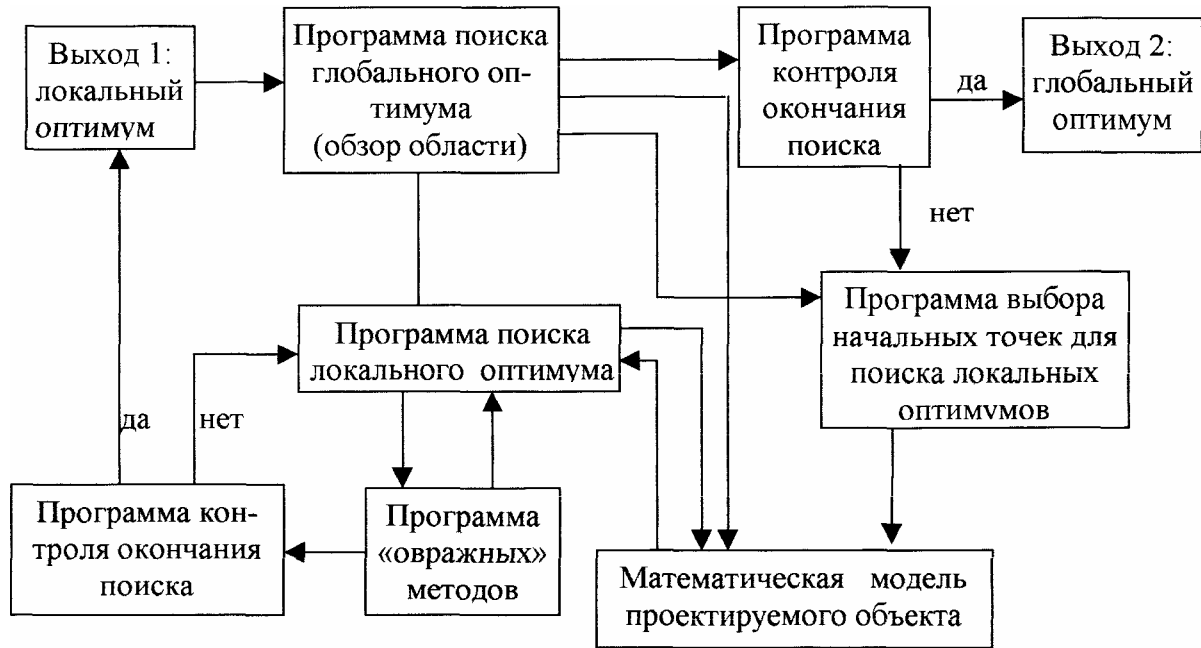


Рис.3.7. Структура программного комплекса оптимального проектирования

Опыт применения программ оптимизации показал, что их подключение к модели параметрического анализа самолета не требует их существенной доработки, тем более, если используются такие современные программные продукты как CATIA, CIMATRON, CADD5 и UNIGRaphics и технический комплекс в виде графстанции типа RS/6000-42T. Если доработки и имеют место, то они связаны в основном с необходимостью нормализовать варьируемые параметры, ограничения и целевую функцию. Нормализация облегчает проблему поиска экстремума, так как при выборе направления и величины шага необходимо оценивать расстояния, то есть нужно вводить ту или иную норму в пространстве параметров. Эта операция требует, чтобы все параметры имели одну размерность или вообще были безразмерными.

Кроме того, нормализация приводит к подобию различные задачи оптимизации, облегчает анализ результатов оптимизации, позволяя сравнивать относительный «вклад» в изменение критерия каждой переменной.

В нормализованном виде все переменные имеют порядок единицы. Рассмотрим следующий алгоритм нормализации:

1. Определить нормализованные переменные x_i ,

$$\overline{x}_i = \frac{x_i - x_i''}{x_i' - x_i''}; \quad 0 \leq \overline{x}_i \leq 1, \quad i = \overline{1, n}. \quad (3.22)$$

2. Определить нормализованные ограничения в виде равенств

$$\overline{\psi}_j(x_i) = \frac{\psi_j(x_i)}{B} = 1, \quad j = \overline{1, m}. \quad (3.23)$$

3. Определить нормализованные ограничения в виде неравенств

$$\overline{E}_k(x_i) \Delta \frac{E_k(x_i)}{C_k^s} \leq 1 \left(\text{или } \overline{E}_k \geq 1 \text{ для } C_k^H \right), \quad k = \overline{1, p}. \quad (3.24)$$

В этих соотношениях символ Δ означает равенство по определению, а индексы «В» и «Н» – верхний и нижний пределы изменения параметров соответственно.

Целевую функцию также можно нормализовать, используя для этого ее значение в начальной точке расчета, задаваемой вектором входных параметров $X_{\text{вх}_i}$ или, что эквивалентно, $\overline{X}_{\text{вх}_i}$. Значение нормализованной целевой функции будет равно

$$\overline{F}(X) \Delta \overline{F}(X) / F(X_{\text{вх}_i}). \quad (3.25)$$

Процедура нормализации требует от проектировщика хорошего понимания физической сути решаемой задачи, знания пределов изменения проектных переменных и порядка значений целевой функции. Обычно это не представляет сложной проблемы для опытного проектировщика. При современных достижениях в области математического программирования может быть решена практически любая задача оптимизации нелинейных систем с несколькими сотнями переменных и ограничений.

3.5. Процедуры выбора критериев оптимальности технологических процессов

При разработке оптимального технологического процесса наиболее важным является обоснование цели и оценка эффективности технологических операций или ее отдельных элементов, например, режимов резания [12].

Под основной целью технологического процесса или в авиастроении (в машиностроении) обычно понимается обеспечение заданных характеристик качества изделия наиболее производительным путем при минимальных затратах. В этом случае **оптимальность операции** можно определить как меру ее соответствия поставленной цели. Чем эффективнее операция, тем выше ее производительность и экономичность. То же можно сказать и о технологическом процессе в целом.

В задачах, которые встречаются в условиях оптимизации технологических процессов (ТП), критерии оптимальности могут быть различными, однако все они должны удовлетворять определенным требованиям:

1. Обладать достаточной полнотой описания объекта;
2. Иметь определенный физический смысл;
3. Быть количественными и выражаться однозначно некоторым числом;
4. Иметь простой математический вид;
5. Определяться с допустимой точностью.

В зависимости от вида и уровня задач оптимизации (расчет режимов резания, проектирование операции и технологического процесса или оценка работы предприятия в целом) основные используемые критерии оптимальности можно подразделить на следующие виды:

1. Стоимостные (экономические): минимальная себестоимость; наименьшие народнохозяйственные приведенные затраты; наименьшие приведенные хозрасчетные затраты; наибольшая прибыль; рентабельность; минимальный уровень затрат на производство (минимальные затраты на электрическую и другие виды энергии, на основные и вспомогательные материалы, на фонд заработной платы и др.).

2. Функциональные «технико-экономические»: максимальная производительность; наименьшее штучное время; основное и вспомогательное время; коэффициент полезного действия оборудования; надежность работы системы оборудования или отдельных ее элементов; станкоемкость изделия; стабильность технологического процесса обработки.

3. Технологические: точность изготовления изделия, показатели качества поверхности изделия (шероховатость, волнистость, микротвердость, остаточные напряжения и др.); физико-химические свойства изделий; стойкость инструмента.

4. Эксплуатационные: износостойкость; усталостная прочность; контактная жесткость и другие показатели долговечности изделий.

5. Прочие: психологические; эстетические, эргономические.

Наибольшее распространение при решении задач оптимизации технологического проектирования получили экономические и технико-экономические критерии оптимальности. Это связано с тем, что в основе разработки любого ТП или решения более частной задачи, например, расчета режимов резания, лежат два принципа: **технический и экономический**. В соответствии с **первым принципом** технологический процесс должен гарантировать выполнение всех требований на изготовление изделия. **Второй принцип** условия, обеспечивающий минимальные затраты труда и наименьшие издержки производства. **Первый принцип** наиболее полно отражается минимальной себестоимостью из группы экономических критериев, а второй - максимальной производительностью из группы технико-экономических критериев.

Технологические и эксплуатационные критерии оптимальности используются при обеспечении требуемого качества наиболее ответственных изделий (точности, качества поверхности, физико-химических свойств и др.), а также эксплуатационных свойств отдельных деталей, определяющих надежность и долговечность машин.

Одним из широко применяемых критериев оптимальности для решения технологических задач и определения режимов резания, в частности, является **максимальная производительность** [7,8].

Производительностью рабочей машины называется количество обрабатываемого продукта за единицу рабочего времени. **Штучная производитель-**

ность (шт./мин) на операции определяется величиной, обратной штучно-калькуляционному времени на эту операцию

$$\Pi = (t_{шт.-к.})^{-1}. \quad (3.26)$$

Как известно, время выполнения операции $t_{шт.-к.}$ состоит из штучного времени $t_{шт}$ и подготовительно-заключительного $T_{п.-з.}$ на всю партию деталей N

$$t_{шт.-к.} = t_{шт} + T_{п.-з.}/N.$$

Составными элементами штучного времени являются: основное t_0 и вспомогательное t_e время, время перерывов на отдых и естественные надобности $t_{отд}$ и время обслуживания рабочего места $t_{обс.}$. Анализ элементов штучного времени $t_{шт}$ показывает, что от режимов резания зависят t_0 (или можно принять машинное t_m) и часть времени $t_{обс.}$, затрачиваемого на смену и подналадку инструмента

$$t_{обс.} = t'_{обс.} + t_{см.}$$

Тогда формула (3.26) примет вид:

$$\Pi = [(t_m + t_{см.}) + (t_e + t'_{обс.} + t_{отд} + T_{п.-з.}/N)]^{-1} = [t_{шт.-к.р.} + t_{шт.-к.н.}]^{-1}, \quad (3.27)$$

где $t_{шт.-к.р.}$ и $t_{шт.-к.н.}$ – часть штучно-калькуляционного времени, соответственно зависящего и независящего от режимов резания.

Таким образом, штучная производительность, зависящая от режимов резания, определяется только величиной

$$t_{шт.-р.} = t_m + t_{см.} \quad (3.28)$$

Машинное время в общем виде

$$t_m = t_p + t_x$$

где t_p – время резания, мин; t_x – время холостого хода, мин.

Тогда, если ввести коэффициент резания $\lambda = L_d/L$, представляющий отношение длины детали L_d к длине рабочего хода L , получим

$$t_m = \lambda t_p. \quad (3.29)$$

Для наиболее распространенных методов обработки металлов резанием (точение, сверление, фрезерование) величина t_p может быть найдена по формуле

$$t_p = \frac{L}{nS} i = \frac{Lh}{nSt}, \quad (3.30)$$

где n – частота вращения заготовки, об/мин;

S – подача, мм/об.;

t – глубина резания, мм;

L – длина обрабатываемой поверхности, мм;

h – величина припусков, мм;

i – число проходов.

Время смены и подналадки инструмента, приведенной к одной детали,

$$t_{см.} = T_{см} t_p / T, \quad (3.31)$$

где $T_{см}$ – время, затрачиваемое на каждую смену инструмента, мин;

T – период стойкости инструмента, мин.

Подставив формулы (3.29), (3.20) и (3.31) в формулу (3.28) и введя обозначение

$$n = \frac{1000v}{\pi D},$$

получим

$$t_{ум\ p} = \frac{\pi DLh}{1000vst} \left(\lambda + \frac{T_{см}}{T} \right). \quad (3.32)$$

Эта формула не может в таком виде использоваться для оптимизации режимов резания, так как в нее входит переменная величина – стойкость инструмента T , зависящая от этих режимов согласно известной формуле для определения скорости резания v

$$v = \frac{C_v}{T^m t^{x_0} S^{y_0}} k_v. \quad (3.33)$$

После подстановки этой зависимости в формулу (3.32) и соответствующих преобразований получим окончательное выражение для критерия оптимальности **минимальное штучное время**

$$t_{ум\ p} = \frac{\pi DLh}{1000vst} \left(\lambda + \frac{T_{см} v^{1/m} t^{x_0/m} S^{y_0/m}}{C_v^m} \right). \quad (3.34)$$

Анализ формулы (4.34) показывает, что в наиболее общем виде этот критерий по отношению к оптимизируемым параметрам v , s и t является нелинейным.

Показатель «**минимальная себестоимость**» охватывает широкий круг затрат общественного труда и наряду с затратами живого труда учитывает затраты прошлого труда, овеществленного в средствах производства (амортизация и ремонт оборудования, энергия, вспомогательные материалы, измерительный инструмент, помещения).

Цеховая себестоимость технологической операции без учета затрат на заготовку определяется следующим образом

$$C_{оп} = C_{з.с.} + C_a + C_{рем} + C_{эн} + C_v + C_{пр} + C_{ин} + C_u,$$

где $C_{з.с.}$ – заработная плата станочников (с начислением в фонд социального страхования);

C_a – амортизационные отношения на замену станка;

$C_{рем}$, $C_{эн}$, C_v , $C_{пр}$, $C_{ин}$, C_u – затраты соответственно на ремонт станка, силовую электроэнергию, вспомогательные материалы, амортизацию и ремонт универсальных приспособлений, амортизацию, ремонт и заточку универсальных режущих инструментов, а также затраты, связанные с использованием помещения.

Процесс обработки заготовки резанием характеризуется двумя факторами: временем $t_{ум}$, требуемым для обработки, и зависящей от него себестоимостью. С изменением режимов обработки эти факторы изменяются, однако простой зависимости между ними не существует. Это связано с тем, что изменение каждого фактора зависит от различных параметров, и в первую очередь от изнашивания инструмента [39,43].

Изменение условий обработки в направлении повышения производительности может привести к настолько большим инструментальным расходам, что они превысят сокращение расходов, достигнутых за счет уменьшения времени обработки. И, наоборот, недостаточное использование режущих способностей инструмента может резко повысить общие расходы (суммарную себестоимость обработки), связанные с временным фактором.

Характер влияния различных затрат на себестоимость операции показан на рис. 3.8. Суммарная себестоимость обработки (кривая 4) включает три вида затрат:

1. Затраты, не зависящие от режимов резания (затраты на вспомогательное время и приходящаяся на заготовку часть подготовительно-заключительного времени $T_{п.-з.} N$);
2. Затраты, пропорциональные времени обработки, которые уменьшаются с сокращением машинного времени (все затраты, которые снижаются с уменьшением t_0 , за исключением затрат на инструмент);
3. Затраты, пропорциональные производительности обработки (инструментальные затраты, зависящие от скорости, подачи, глубины резания, а также инструментального и обрабатываемого материалов).

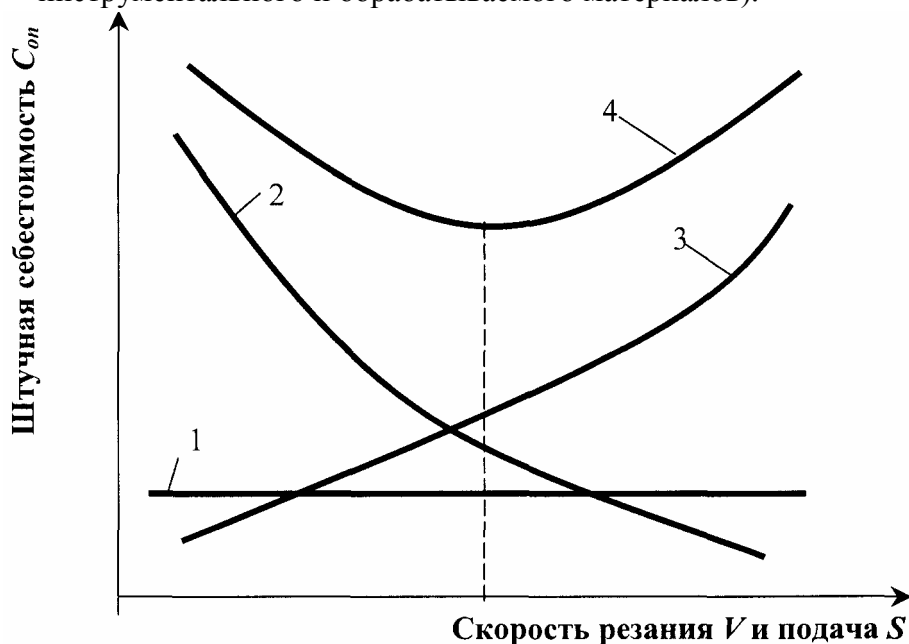


Рис. 3.8. Зависимости себестоимости обработки от скорости резания V и подачи S

Кривая суммарной себестоимости обработки имеет минимум, положение которого зависит от характера кривых 2 и 3 (рис.3.8). В большинстве случаев эта кривая имеет характер гиперболы, зависящей от стоимости станко-минуты и стоимости амортизации рабочего места [41, 44].

При определении технологической себестоимости операции могут использоваться бухгалтерский метод расчета, метод определения себестоимости

станко-часа и метод поэлементного расчета. Последний является наиболее точным методом и используется в дальнейших расчетах. Элементы технологической себестоимости можно условно разделить на две группы, одна из которых не зависит ($C_{рем}, C_{в}$), а вторая зависит от режимов резания ($C_{зс}, C_a, C_{пр}, C_{ин}, C_{эн}, C_n$). Вторая группа элементов себестоимости операции пропорциональна штучному времени, поэтому ее целесообразно привести к 1 мин работы оборудования ($C'_{зс}, C'_a, C'_{пр}, C'_{эн}, C'_n$).

Так, затраты на заработную плату станочника [25]

$$C_{зс} = Z_{мин} K_T t_{шт п},$$

где $Z_{мин}$ – тарифная ставка 1-го разряда данной категории рабочих, кол./мин;

K_T – тарифный коэффициент;

$t_{шт п}$ – часть штучно-калькуляционного времени, зависящая от режимов резания, мин.

Произведя замену в рассмотренной формуле, можно получить

$$C_{зс} = C'_{зс} t_{шт п}.$$

Аналогично могут быть найдены затраты, определяющие элементы себестоимости по амортизации станка C_a , эксплуатации приспособлений $C_{пр}$, амортизации помещений C_n и силовой электроэнергии $C_{эн}$, приведенных к 1 минуте работы оборудования

$$C_a = C'_a t_{шт п}; \quad C_{пр} = C'_{пр} t_{шт п}; \quad C_n = C'_n t_{шт п}; \quad C_{эн} = C'_{эн} t_{шт п}.$$

В этом случае себестоимость операции, зависящая от режимов резания, может быть определена по зависимости

$$C_{оп пер} = C'_{зс} t_{шт п} + C'_a t_{шт п} + C'_{пр} t_{шт п} + C'_{эн} t_{шт п} + C'_n t_{шт п} + C_{ин},$$

где $C_{ин}$ – инструментальные расходы, приведенные к одной заготовке.

Введя обозначение $R = C'_{зс} + C'_a + C'_{пр} + C'_{эн} + C'_n$, определим положение минимума кривой штучной себестоимости $C_{оп пер}$ (рис.3.8) расчетным путем, используя выражение

$$C_{оп пер} = R t_{шт п} + C_{ин}. \quad (3.35)$$

Инструментальные расходы на одну деталь можно определить по формуле

$$C_{ин} = M/G,$$

где M – инструментальные расходы, приведенные к одному периоду стойкости,

G – число деталей, обрабатываемых за период стойкости.

$$\text{При этом } q = \frac{T}{t_p},$$

где t_p – время резания инструментом каждой заготовки, мин.

Инструментальные расходы, приведенные к одному периоду стойкости

$$M = \frac{S_{ин}}{n_T} + C'_{пер} + C'_{зам ин},$$

где S – покупная (начальная) стоимость инструмента,

$C'_{пер}$ - стоимость переточки инструмента, приведенная к одному периоду стойкости;

n_1 - число периодов стойкости. После подстановки всех зависимостей

$$C_{он пер} = Rt_{шт п} + \frac{Mt_p}{T}. \quad (3.36)$$

Учитывая, что $t_{шт п} = t_m + t_{см}$, а $t_m = t_p \lambda$, и введя зависимость (3.30) для времени резания t_p , получим

$$C_{он пер} = \frac{Lh}{nst} \lambda R + \frac{T_{см}}{T} \frac{LhR}{nst} + \frac{1}{T} M \frac{Lh}{nst}. \quad (3.37)$$

Для получения окончательного вида оценочной функции «минимальная себестоимость» производится замена переменных n и T их зависимостями от скорости резания v , и после несложных преобразований имеем

$$C_{он пер} = \frac{\pi DLh}{10^3 vst} \left[\lambda R + \frac{(RT_{см} + M)v^{1/m} t^{x/m} S^{y/m}}{vst} \right]. \quad (3.38)$$

Как видно из формулы (4.38), вид критерия оптимальности по параметрам v , s и t является нелинейным. Это обстоятельство затрудняет разработку методов оптимизации, поэтому в частных случаях стремятся упростить вид критерия оптимальности, что естественно, вносит определенные погрешности в получаемые результаты. Таким примером является широко используемый метод линейного программирования для расчета оптимальных режимов резания v и s . В этом случае критерий «минимальная себестоимость» рассматривается в виде

$$C_{он пер} = C_1 / ns,$$

где C_1 - некоторая постоянная, характеризующая условие обработки.

При решении задач оптимизации процессов механической обработки часто возникает необходимость одновременного достижения нескольких противоречащих друг другу целей. Принимая решения, улучшающие оценки одного критерия, например, минимальной себестоимости операции, ухудшаются тем самым оценки по другим критериям, например, наибольшей производительности и др. В таких случаях возникает задача оценки и сравнения различных проектных решений при так называемом **векторном критерии эффективности**. С этой целью используют обобщенные критерии, которые являются скалярными функциями частных критериев и учитывают степень достижения всех целей в совокупности, отражая относительную значимость каждого критерия в отдельности.

Поскольку каждый из частных критериев является фактически функцией управляемых переменных, то и обобщенный критерий в свою очередь можно рассматривать как некоторую функцию управляемых переменных. Эту функцию, как отмечалось выше, обычно называют **целевой**. При таком подходе, называемом **свертыванием векторного критерия**, задача сравнения решений по

векторному критерию фактически заменяется задачей выбора способа свертывания и определения значения коэффициентов, участвующих в этом свертывании. Существуют следующие виды обобщенных (свернутых) критериев.

Аддитивный критерий. В этом случае в качестве обобщенного критерия берется «взвешенная» сумма частных критериев

$$F(X) = \sum_{j=1}^I \alpha_j k_j(x), \quad (3.39)$$

где x – управляемые переменные;

α_j – неотрицательные коэффициенты, значения которых выбираются, исходя из степени важности отдельных целей, и определяются на основании ранее решенных аналогичных задач или методом проб. В последнем случае «весовые» коэффициенты подбираются при анализе результатов, получаемых при различных значениях этих коэффициентов, α_j является неуправляемой переменной.

Мультипликативный критерий. В отличие от предыдущего в данном случае в качестве обобщенного критерия берется «взвешенное» произведение частных критериев

$$F(x) = \prod_{j=1}^I [K_j(x)]^{\alpha_j}. \quad (3.40)$$

Конъюнктивный критерий. По этому критерию оценивается каждое решение с точки зрения цели, степень достижения которой (с учетом «весового» коэффициента) в данном случае наименьшая

$$F(x) = \min_{1 \leq j \leq J} \alpha_j k_j(x). \quad (3.41)$$

Дизъюнктивный критерий. Противоположен предыдущему и оценивает решение с точки зрения цели, степень достижения которой (также с учетом «весового» коэффициента) в данном случае максимальна

$$F(x) = \max_{1 \leq j \leq J} \alpha_j k_j(x). \quad (3.42)$$

Могут использоваться и другие способы свертывания критериев. Например, способ выделения наиболее важного критерия предусматривает определение из набора частных критериев оптимальности одного, который принимается за обобщенный критерий допустимости. Для выбора наиболее важного критерия может быть рекомендован **метод последовательных уступок**, предусматривающий упорядочивание всех количественных целей в порядке убывания их значимости.

3.6. Моделирующий алгоритм для анализа оптимальности системы массового обслуживания

Наличие имитационной модели, реализуемой на вычислительной машине, позволяет провести интересные в теоретическом и практическом отношении исследования системы массового обслуживания, к которым относится лю-

бая большая система автоматизированного проектирования конструкторско-технологических разработок.

Большой интерес представляет исследование влияния вариаций параметров системы на показатели, характеризующие ее основные свойства.

На этом пути могут быть получены рекомендации, полезные с точки зрения синтеза системы. Важнейшим этапом такого исследования можно считать оптимизацию параметров системы, в основу которой положены показатели эффективности.

Рассмотрим трехканальную систему массового обслуживания (система автоматизированного проектирования: оснастки, технологических процессов и разработки управляющих программ для станков с ЧПУ) с продолжительностями обслуживания, имеющими показательное распределение с параметрами μ_1 , μ_2 и μ_3 соответственно номерам каналов.

Заявки поступают в систему с двумя пуассоновскими потоками с интенсивностями λ_1 и λ_2 . Порядок обслуживания поступивших заявок состоит в следующем. Заявка первого потока (например, заявка на проектирование какого-либо вида приспособления для мехобработки) поступает на обслуживание в систему на первый канал и, если он занят – на второй. В случае занятости второго канала заявка получает отказ. Заявка второго потока поступает на второй канал; если он занят – то на третий, а если и последний занят – получает отказ.

Отказ в обслуживании заявки первого потока приводит к убытку α_1 , а второго – к убытку α_2 . Если заявки отсутствуют, то простой в системе первого канала в единицу времени приводит к убытку C_1 , второго канала – к убытку C_2 и третьего – к убытку C_3 .

В связи с упомянутой системой массового обслуживания нас будет интересовать выбор оптимальных значений средней длительности обслуживания μ_1 , μ_2 и μ_3 . Здесь представляют интерес два случая:

1. Выбор величины μ_i при отсутствии ограничений на их сумму (задача **на безусловный экстремум**);
2. Выбор величин μ_i , когда сумма их ограничена (задача **на условный экстремум**).

Рассмотрим сначала первую задачу. Требуется определить такие μ_1 , μ_2 и μ_3 , чтобы суммарный убыток при эксплуатации системы в течение заданного времени $(0, T)$ был минимальным.

Заметим, что такая задача (безусловный экстремум) имеет точное решение, полученное при аналитическом описании процесса функционирования системы.

Построение моделирующего алгоритма для данной системы массового обслуживания не должно представлять затруднений.

Формируется поток однородных событий с показательным распределением интервалов времени между событиями, параметром которого служит

$\lambda = \lambda_1 + \lambda_2$. Из этого (суммарного) потока по жребью с вероятностью $\lambda_1/(\lambda_1 + \lambda_2)$ выделяются события, которые принадлежат первому потоку.

Остальные события принадлежат второму потоку.

Далее, как обычно, проверяется занятость каналов и возможность обслужить ту или другую заявки. Если такая возможность имеется, для соответствующего канала формируется время обслуживания τ по показательному закону с параметром μ_i , где номер i совпадает с номером обслуживающего канала.

Обратим внимание на порядок формирования величины убытка (U) при моделировании процесса функционирования рассматриваемой системы автоматизированного проектирования

$$U = \alpha_1 m_1 + \alpha_2 m_2 + \mu_1 \Sigma_1 + \mu_2 \Sigma_2 + \mu_3 \Sigma_3, \quad (3.43)$$

где m_1 и m_2 – средние значения количества отказов для заявок первого и второго потоков, а $\Sigma_1, \Sigma_2, \Sigma_3$ – средние суммарные времена простоев для соответствующих каналов.

Сказанного достаточно для представления о работе этого весьма простого моделирующего алгоритма.

Остановимся кратко на процедуре оптимизации. Выбираем точку нулевого приближения $\mu_1^0, \mu_2^0, \mu_3^0$. Затем, обращаясь к моделирующему алгоритму для исследуемой системы массового обслуживания (автоматизированной системы проектирования конструкторско-технологических разработок), при $\mu_1 = \mu_1^0, \mu_2 = \mu_2^0, \mu_3 = \mu_3^0$ проводим моделирование процесса функционирования системы в интервале времени $(0, T)$ с количеством реализаций модели, требуемым заданной точностью решения задачи. В результате моделирования получаем значения величин $m_1, m_2, \Sigma_1, \Sigma_2$ и Σ_3 , а затем по формуле (3.43) вычисляем убыток $U^{(0)}$.

Далее последовательно для каждого переменного (μ_1, μ_2, μ_3) проводим минимизацию убытка как функции одной переменной, фиксируя значения других переменных и обращаясь каждый раз к модели системы массового обслуживания для определения убытка. Таким образом, получаем точку первого приближения $\mu_1^{(1)}, \mu_2^{(1)}, \mu_3^{(1)}$ и соответствующее значение убытка $U^{(1)}$. Аналогичным путем, исходя из точки первого приближения, можно получить второе приближение и т.д.

В качестве примера [21] приведем результаты восьми приближений для рассматриваемой задачи при $\lambda_1=0.4$ и $\lambda_2=0.6$:

$$\begin{aligned} \mu_1^{(0)} &= 1, & \mu_3^{(0)} &= 4, \\ \mu_2^{(0)} &= 3, & U^{(0)} &= 35525, & \mu_1^{(1)} &= 5, \\ \mu_2^{(1)} &= 8.75, & U^{(1)} &= 34257; & \mu_1^{(5)} &= 7.5, \\ \mu_3^{(1)} &= 7.5, & & & \mu_2^{(5)} &= 7.5, & U^{(5)} &= 29544; \\ \mu_1^{(2)} &= 8.12, & & & \mu_3^{(5)} &= 5, \\ \mu_2^{(2)} &= 5, & U^{(2)} &= 34009; & \mu_1^{(6)} &= 5, \end{aligned}$$

$$\begin{aligned}
\mu_3^{(2)} &= 9.84, & \mu_2^{(6)} &= 7.5, \quad U^{(6)} = 29377; \\
\mu_1^{(3)} &= 7.5, & \mu_3^{(6)} &= 1.72, \\
\mu_2^{(3)} &= 7.5, \quad U^{(3)} = 29795, & \mu_1^{(7)} &= 5, \\
\mu_3^{(3)} &= 8.75, & \mu_2^{(7)} &= 8.75, \quad U^{(7)} = 29356,9; \\
\mu_1^{(4)} &= 8.75, & \mu_3^{(7)} &= 5, \\
\mu_2^{(4)} &= 9.53, \quad U^{(4)} = 29697; & \mu_1^{(8)} &= 5, \\
\mu_3^{(4)} &= 8.12, & \mu_2^{(8)} &= 8.75, \quad U^{(8)} = 29356.9, \\
& & \mu_3^{(8)} &= 5.
\end{aligned}$$

По-видимому, результаты седьмого и восьмого приближений могут отличаться лишь в пределах точности вычислений.

В случае, если заданы дополнительные ограничения на величины μ_1, μ_2 и μ_3 , например, функции вида g_h в выражении

$$h = 1, 2, \dots, h^*, \quad (3.44)$$

получаем задачу на **условный экстремум**. Одним из условий такого рода может быть условие

$$\mu_1 + \mu_2 + \mu_3 \leq p,$$

которое можно понимать в некотором смысле как ограничение суммарных возможностей обслуживающей системы.

В этом случае рассмотренная выше процедура минимизации должна применяться к выражению

$$f = U + \frac{1}{C} l^{c(\mu_1 + \mu_2 + \mu_3 - p)},$$

где величина (U) задается соотношением (3.43)

Таким образом, рассмотренная трехканальная система массового обслуживания с примером и процедуры оптимизации этой системы на основе моделирующего алгоритма дают основания полагать, что расчеты в виде математической модели выглядят не сложно, но убедительно демонстрируют правильность выбора технического решения, в основу которого положены показатели эффективности принимаемых технических решений.

4. ПРОЦЕДУРЫ ПРИНЯТИЯ ОПТИМАЛЬНОГО РЕШЕНИЯ ПРИ ВЫЧИСЛЕНИИ ЭКСТРЕМУМОВ ДЛЯ АНАЛИЗА СИСТЕМЫ

Задачи современной теории систем просто формируются в виде задач на экстремум. Их оптимальное решение сводится к поиску экстремума, что является нормализованным методом, когда необходимо найти значение параметров, минимизирующих или максимизирующих зависящую от них величину. Для решения задач на экстремум с ограничением процедур принятия оптимального решения используется метод множителей Лагранжа, а также методы линейного и нелинейного программирования, вариационного исчисления принципа максимума. Рассмотрим некоторые вопросы принятия оптимального решения, с использованием вышеназванных методов и процедур.

4.1. Экстремум без ограничений как процедура принятия оптимального решения

Пусть D – подмножество (оператор действия), принадлежащее нормированному линейному векторному пространству, с нормой $\left\| \begin{smallmatrix} \dots \\ \dots \end{smallmatrix} \right\|$ и пусть f – действительная функция, определенная на D , то есть $f: D \rightarrow R$, где R^n – евклидово n -мерное пространство и $R = R^1$, а функция $\left\| \begin{smallmatrix} \dots \\ \dots \end{smallmatrix} \right\|: x \rightarrow R$ называется нормой, (где X является линейным пространством), если она удовлетворяет следующим условиям:

1. $\|x\| \geq 0$, для всех $x \in X$;
2. $\|x\| = 0$, если только $x = 0$;
3. $\|x + y\| \leq \|x\| + \|y\|$, для всех $x, y \in X$;
4. $\|\alpha x\| = |\alpha| \|x\|$, для всех $\alpha \in R, x \in X$.

Тогда можно показать, что функция $\left\| \begin{smallmatrix} \dots \\ \dots \end{smallmatrix} \right\|$ непрерывна и удовлетворяет неравенству

$$\|x\| - \|y\| \leq \|x - y\|$$

Функция f имеет относительный максимум (минимум) в точке $\alpha \in D$ тогда, если существует положительное действительное число δ , такое, что для

любого $x \in D$ выполняется условие, $\|x - \alpha\| < \delta$, $f(x) \leq f(\alpha)$ [$f(\alpha) \leq f(x)$]. Если эти неравенства выполняются для всех $x \in D$, то f имеет абсолютный минимум или максимум.

Точка $\alpha \in D$ является внутренней точкой D тогда, если существует такое положительное действительное число δ , что если x удовлетворяет условию $\|x - \alpha\| < \delta$, то $x \in D$. Пусть D – действительная ось, то есть $D = \mathbb{R}$, и предположим, что существует производная $df(x)/dx$, непрерывная в точке $x = \alpha$. Тогда необходимым условием того, что точка α является внутренним максимумом или минимумом, будет

$$\left. \frac{df(x)}{dx} \right|_{x=\alpha} = 0 \quad (4.1)$$

Предположим, что f имеет непрерывную вторую производную в точке $x = \alpha$. Известно, что если удовлетворяется уравнение (4.1) и если $[d^2f(x)/dx^2]_{x=\alpha}$ положительно (отрицательно), то α является отрицательным минимумом (максимумом). Минимумы и максимумы изобразим графически (рис. 4.1).

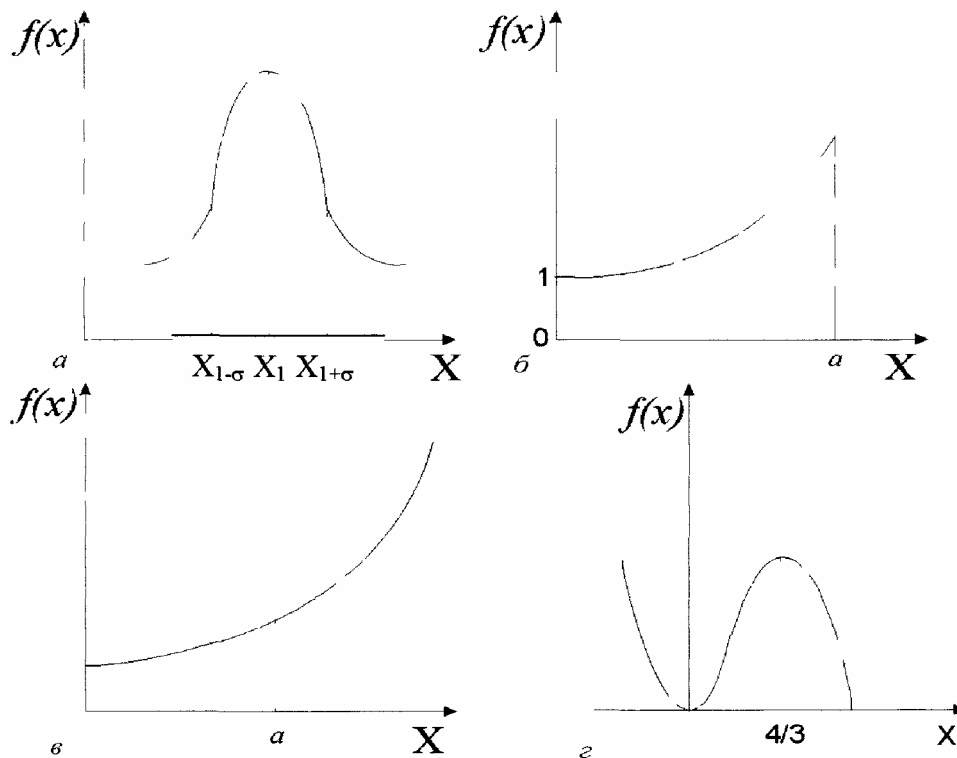


Рис. 4.1. Примеры экстремумов при определении относительного решения.

Метод нахождения экстремумов может быть обобщен на функции более чем одной скалярной переменной благодаря использованию частных производных. Пусть $f: R^n \rightarrow R$. Аналогично уравнению (4.1) необходимым условием того, что вектор α является максимумом или минимумом, будет

$$\left. \frac{\partial f(x)}{\partial x} \right|_{x=\alpha} = 0, \quad (4.2)$$

причем существуют все, частные производные, непрерывные при $x = \alpha$.

Здесь производится дифференцирование по векторному параметру, то есть $f = f(x)$ – скалярная функция. Тогда дифференцирование относительно вектора x определим следующим образом:

$$\frac{df}{dx} = \left[\frac{df}{dx_1}, \frac{df}{dx_2}, \dots, \frac{df}{dx_n} \right]^T$$

Часто эта производная называется **градиентом** и обозначается символом $\nabla_x f = df/dx$.

Если $z = z(x)$, то:

$$\frac{dz^T}{dX} = \begin{bmatrix} \frac{dz_1}{dx_1}, \frac{dz_2}{dx_1}, \dots, \frac{dz_m}{dx_1} \\ \frac{dz_1}{dx_2}, \frac{dz_2}{dx_2}, \dots, \frac{dz_{m_l}}{dx_2} \\ \vdots \\ \frac{dz_1}{dx_n}, \frac{dz_2}{dx_n}, \dots, \frac{dz_m}{dx_n} \end{bmatrix}$$

В некоторых источниках [13] эта матрица называется якобианом и обозначается символом $J_{\dot{x}z}(x)$, где

$$\left[\frac{dz}{dx} \right]_{ij} = [J_x \{z(X)\}]_{ij} = \left[\frac{\partial z_i}{\partial x_j} \right].$$

Предположим, что существуют вторые частные производные непрерывные, при $x=\alpha$, тогда достаточным условием того, чтобы α , была относительным максимумом (минимумом), является справедливость уравнения (4.2) и отрицательность (положительность) определения, предоставляющего собой матрицу, общий элемент которой равен $\left[\partial^2 f(x) / \partial x_i \partial x_j \right]_{x=\alpha}$.

Квадратичную форму относительного X введем с помощью равенства

$$Q = X^T F x = \sum_{i=1}^n \sum_{j=1}^n f_{ij} x_i x_j, \quad (4.3.)$$

где F – квадратная симметричная матрица. Квадратичная форма называется положительно определенной, если $Q > 0$ при $X \neq 0$. Квадратичная симметричная матрица называется положительно определенной, если положительно определена соответствующая ей квадратичная форма. Необходимым и достаточным условием положительной определенности матрицы F является выполнение следующих неравенств:

$$|f_{11}| > 0, \quad \begin{vmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{vmatrix} > 0, \quad \begin{vmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{vmatrix} > 0 \quad \text{и т.д.}$$

Если в этих неравенствах знак «>» заменить знаком « $\geq$ », то соответствующая матрица называется **положительно полуопределенной**. Матрица F называется **отрицательно определенной** или **отрицательно полуопределенной**, если умножение ее на (-1) приводит к положительно определенной или положительно полуопределенной матрице.

Рассмотрим процедуру вычисления **экстремума** для принятия оптимального решения на примере.

Найдем максимум функции

$$f(x) = \frac{1}{(x_1 - 1)^2 + (x_2 - 1)^2 + 1}, \quad (4.4)$$

где $X^T = [x_1, x_2]$, а X^T – используется для обозначения **транспортирования** вектора – столбца. В соответствии с обобщенным вариантом изложенной выше скалярной процедуры запишем частные производные от f по x_1 и x_2 , приравняем их к нулю, то есть

$$\begin{aligned} \frac{\partial f}{\partial x_1} &= \frac{(-1)(2x_1 - 2)}{[(x_1 - 1)^2 + (x_2 - 1)^2 + 1]^2} = 0, \quad \alpha_1 = 1, \\ \frac{\partial f}{\partial x_2} &= \frac{(-1)(2x_2 - 2)}{[(x_1 - 1)^2 + (x_2 - 1)^2 + 1]^2} = 0, \quad \alpha_2 = 1. \end{aligned}$$

Таким образом, поскольку $\alpha_1 = \alpha_2 = 1$ – единственный экстремум, и что при проверке результатов вычислений вторые производные в точке экстремума отрицательные, видно, что максимум находится в точке $x^T = [1, 1]$. Изобразим функцию $f(x)$: $f(x) = 1/[(x_1 - 1)^2 + (x_2 - 1)^2]$ графически (рис. 4.2).

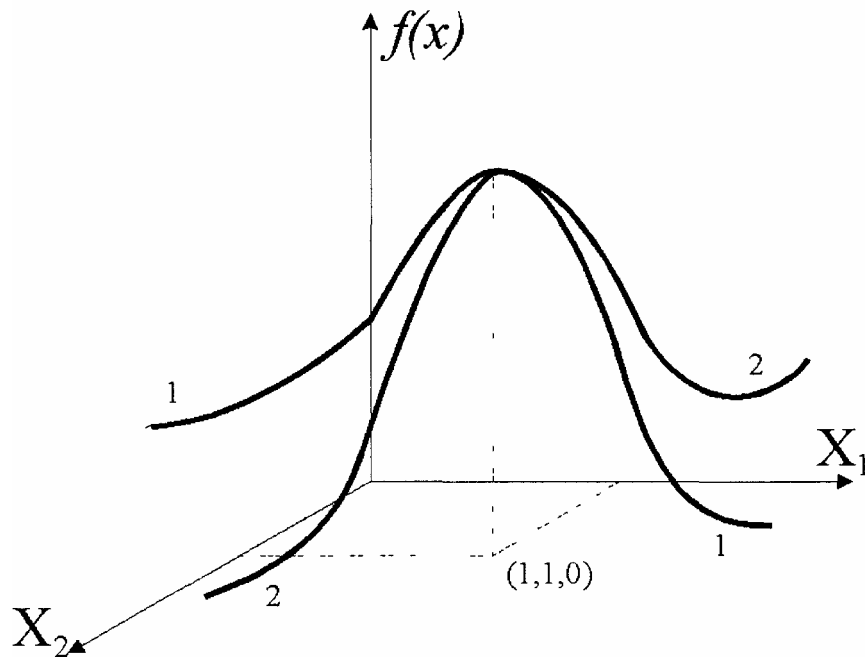


Рис 4.2 Экстремум функции $f(x) = 1/[(x_1-1)^2 + (x_2-1)^2]$

4.2. Анализ оптимального решения системы при вычислении экстремума с ограничениями в виде равенств

Исследуем экстремум функции ${}^oF: R^n \rightarrow R$, на которую наложено ограничение в виде равенства $f(x) = 0$, где $f: R^n \rightarrow R^m$. Для решения этой задачи можно использовать несколько методов, два из которых – метод **прямой подстановки** и метод **множителей Лагранжа**. Прямая подстановка – это наиболее оптимальная процедура, однако часто она оказывается более громоздкой, чем метод множителей Лагранжа. Для простоты иллюстраций положим $n = 2$ и $m = 1$.

Метод прямой подстановки может быть описан следующим образом. Допустим, что f такая функция, что существует некоторое $h: R \rightarrow R$, удовлетворяющее равенству $h(x_2) = x_1$. Далее, ${}^oF(x_1, x_2) = {}^oF[h(x_2), x_2]$ является функцией только x_2 и можно применять методы, используемые в задачах без ограничений. Рассмотрим вышестоящее исследование на примере **экстремизации** функции, на которую наложено ограничение, заключающееся в том, что **евклидова норма** X равна единице, то есть $\|X\| = 1$. Математически квадрат евклидовой нормы можно записать в виде $\|X\|^2 = x_1^2 + x_2^2 + \dots + x_n^2 = \langle X, X \rangle$. Поскольку размерность в рассматриваемом примере равна двум, квадрат евклидовой нормы равен $\|X\|^2 = x_1^2 + x_2^2$.

Один из методов решения, этой задачи заключается в том, чтобы выразить x_1 через x_2 , а затем записать ${}^oF = f(X)$ только через x_2 . Теперь можно использовать стандартную скалярную процедуру. Исходя из заданного ограничения на

длину евклидовой нормы, получим $x_1 = (1 - x_2^2)^{1/2}$. Подставив это выражение в ${}^oF(x)$ (в формулу рис. 4.2), получим:

$${}^oF(x_2) = \frac{1}{[(\pm\sqrt{1-x_2^2} - 1)^2 + (x_2 - 1)^2 + 1]}, \quad (4.5)$$

где заданные ограничения входят в ${}^oF(x_2)$. Следует этап – дифференцирование этого выражения по оставшейся переменной x_2 и приравнивание результата нулю. При этом будут получены два решения. Исследование второй производной показывает, что максимум, который можно показать, что он является абсолютным максимумом, соответствует $x^T = [0,707, 0,707]$, а абсолютный минимум $x^T = [-0,707, -0,707]$. Изобразим это графически (рис. 4.3).

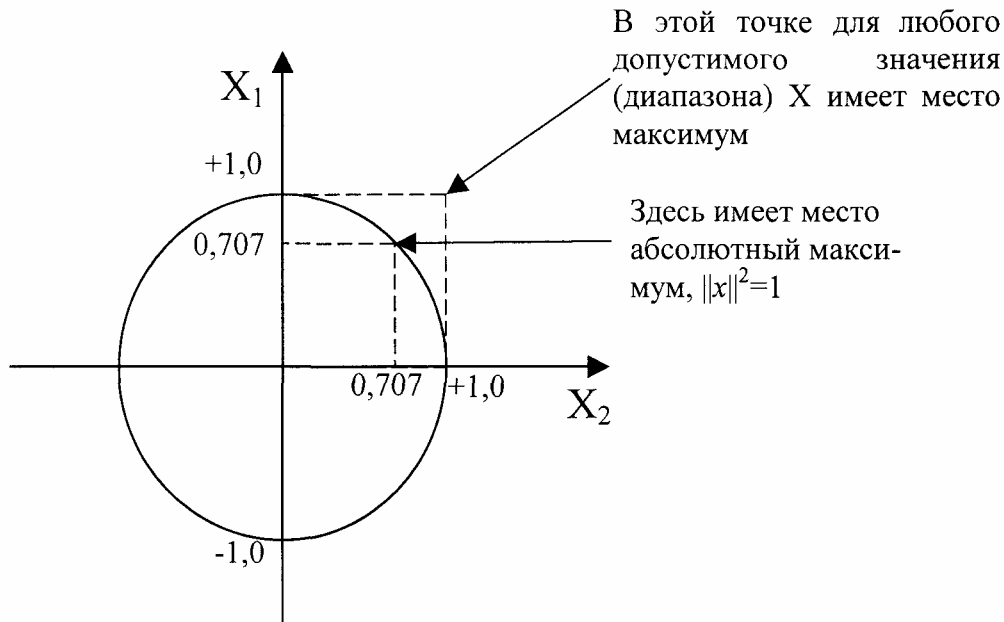


Рис. 4.3. Вид изображения на рис. 4.2. сверху – область пространства состояний, определяемая выражением $\|x\|^2 < 1$.

Здесь следует отметить, что в данной процедуре решения задачи, когда нет ограничений в форме **равенств**, относительный минимум отсутствует.

Далее, рассмотрим метод множителей Лагранжа. Предположив, что все соответствующие частные производные существуют, получим:

$$\frac{d{}^oF(x)}{dx_1} = \frac{\partial {}^oF(x)}{\partial x_1} + \frac{\partial {}^oF(x)}{\partial x_2} \frac{dx_2}{dx_1} \quad (4.6)$$

и

$$\frac{df(x)}{dx_1} = \frac{\partial f(x)}{\partial x_1} + \frac{\partial f(x)}{\partial x_2} \frac{dx_2}{dx_1}. \quad (4.7)$$

Поскольку на X должно быть наложено ограничение из условия $f(x)=0$, то из уравнения (4.7) следует, что ограниченная экстремаль должна удовлетворять условию

$$\frac{df(x)}{dx_1} + \frac{\partial f(x)}{\partial x_2} \frac{dx_2}{dx_1} = 0.$$

Из (4.6.) следует, что экстремаль, кроме того, должна удовлетворять условию

$$\frac{\partial^0 F(x)}{\partial x_1} + \frac{\partial^0 F(x)}{\partial x_2} \frac{dx_2}{dx_1} = 0.$$

Предположив, что все приведенные выше частные производные не равны нулю, получим

$$\frac{dx_2}{dx_1} = \frac{-\partial^0 F(x)}{\partial x_1} \bigg/ \frac{\partial^0 F(x)}{\partial x_2} = \frac{-\partial f(x)}{\partial x_1} \bigg/ \frac{\partial f(x)}{\partial x_2}, \quad (4.8)$$

то есть отношения

$$\frac{\partial^0 F(x)}{\partial x_1} : \frac{\partial^0 F(x)}{\partial x_2} \quad \text{и} \quad \frac{\partial f(x)}{\partial x_1} : \frac{\partial f(x)}{\partial x_2}$$

должны быть пропорциональны вектору, предполагаемому в качестве внутренней экстремали. Пусть λ - коэффициент пропорциональности, который называют множителем Лагранжа. Таким образом,

$$\lambda = \frac{-\partial^0 F(x)}{\partial x_1} \bigg/ \frac{\partial^0 F(x)}{\partial x_2} = \frac{-\partial f(x)}{\partial x_1} \bigg/ \frac{\partial f(x)}{\partial x_2} \quad (4.9)$$

Определить функцию Лагранжа

$$L(x, \lambda) = {}^0F(x) + \lambda f(x).$$

Тогда (4.9) эквивалентно $\partial L / \partial x = 0$, а $f(x) = 0$ эквивалентно $\partial L / \partial \lambda = 0$.

Метод множителей Лагранжа можно обобщить на случай произвольных, конечных, положительных, целых n и m , но теперь $\lambda \in R^m$. Если α соответствует условиям максимума или минимума, то имеется $\lambda \in R^m$ такое, что при α и λ , $\partial L / \partial x = 0$ и $\partial L / \partial \lambda = 0$.

Для большинства системных задач необходимо различать вектор **управления** и вектор **состояния**. Если имеется процесс с многими состояниями или непрерывный процесс, то вектор управления $u(k)$ и $u(t)$ определяется из условия минимума или максимума некоторого скалярного показателя характеристики системы. Этот показатель обозначим через J . Вероятнее всего, простая процедура принятия оптимального решения с ограничениями в форме равенств должна минимизировать или максимизировать скалярный показатель характеристики

$$J = {}^0F[x, u] \quad (4.10)$$

при ограничении

$$f(x, u) = 0, \quad (4.11)$$

где $(x - n)$ – мерный вектор

$$\mathbf{x}^T = [x_1, x_2, \dots, x_n],$$

$(u - m)$ – мерный вектор

$$\mathbf{U}^T = [u_1, u_2, \dots, u_m],$$

$(f - n)$ – мерная векторная функция

$$\mathbf{f}^T(\mathbf{x}, \mathbf{u}) = [f_1(\mathbf{x}, \mathbf{u}), f_2(\mathbf{x}, \mathbf{u}), \dots, f_n(\mathbf{x}, \mathbf{u})].$$

Решение находим следующим образом. Объединяем (4.10) и (4.11) множителем Лагранжа, в результате чего получается скалярная величина

$$L(\mathbf{x}, \mathbf{u}, \boldsymbol{\lambda}) = {}^0F(\mathbf{x}, \mathbf{u}) + \boldsymbol{\lambda}^T \mathbf{f}(\mathbf{x}, \mathbf{u}),$$

$$\boldsymbol{\lambda}^T = [\lambda_1, \lambda_2, \dots, \lambda_{n_1}].$$

Далее, находим $\mathbf{x}$ и $\mathbf{u}$ для которых L принимает максимальное или минимальное значение. Для этого:

$$\frac{\partial L}{\partial \mathbf{x}} = \frac{\partial {}^0F}{\partial \mathbf{x}} + \frac{\partial}{\partial \mathbf{x}} \mathbf{f}^T(\mathbf{x}, \mathbf{u}) \boldsymbol{\lambda} = 0,$$

$$\frac{\partial L}{\partial \mathbf{u}} = \frac{\partial {}^0F}{\partial \mathbf{u}} + \frac{\partial}{\partial \mathbf{u}} \mathbf{f}^T(\mathbf{x}, \mathbf{u}) = 0,$$

$$\text{где} \quad \left[\frac{\partial L}{\partial \mathbf{u}} \right]^T = \left[\frac{\partial L}{\partial u_1}, \frac{\partial L}{\partial u_2}, \dots, \frac{\partial L}{\partial u_m} \right].$$

Таким образом, $\partial L / \partial \mathbf{u}$ можно интегрировать как градиент L по $\mathbf{u}$, что обычно обозначается $\nabla_{\mathbf{u}} L$. Также

$$\frac{\partial}{\partial \mathbf{x}} \mathbf{f}^T(\mathbf{x}, \mathbf{u}) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_2}{\partial x_1} & \dots & \frac{\partial f_n}{\partial x_1} \\ \dots & \dots & \dots & \dots \\ \frac{\partial f_1}{\partial x_n} & \frac{\partial f_2}{\partial x_n} & \dots & \frac{\partial f_n}{\partial x_n} \end{bmatrix}. \quad (4.12)$$

Следует заметить, что уравнение (4.12) сходно с транспонированным вектора Якоби

$$[J_x f(x, u)]^T = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_2}{\partial x_1} & \dots & \frac{\partial f_n}{\partial x_1} \\ \frac{\partial f_1}{\partial x_2} & \frac{\partial f_2}{\partial x_2} & \dots & \frac{\partial f_n}{\partial x_2} \\ \dots & \dots & \dots & \dots \\ \frac{\partial f_1}{\partial x_n} & \frac{\partial f_2}{\partial x_n} & \dots & \frac{\partial f_n}{\partial x_n} \end{bmatrix},$$

где $\partial f(x, u)/\partial u$ матрица, а не детерминант.

Для того чтобы величина J достигла экстремума, необходимо, чтобы не только

$$\frac{\partial L}{\partial x} = 0, \frac{\partial L}{\partial u} = 0,$$

но также, чтобы вторая вариация L была больше нуля в случае минимума или меньше нуля в случае максимума. Для того чтобы показать, что означает это ограничение и используя необходимые условия, требующиеся для выполнения, чтобы $J(x, u)$ имела экстремум, найдем вторую вариацию $L(x, u, \lambda)$. Первая вариация $L(x, u, \lambda)$ равна

$$\delta L = \left(\frac{\partial L}{\partial x} \right)^T \delta x + \left(\frac{\partial L}{\partial u} \right)^T \delta u,$$

что представляет собой линейную часть приращения

$$\Delta L = L[X + \delta x, u + \delta u] - L[x, u]. \quad (4.13)$$

Для получения второй вариации L , обозначаемой $\delta^2 L$, возьмем член второго порядка в разложении (4.13) в ряд Тейлора в окрестности $\delta u = 0$, $\delta x = 0$. Получим

$$\delta^2 L = \frac{1}{2} \delta x^T \left\{ \left[\frac{\partial}{\partial x} \frac{\partial L}{\partial x} \right] \delta x + \left[\frac{\partial}{\partial u} \frac{\partial L}{\partial x} \right] \delta u \right\} + \frac{1}{2} \delta u^T \left\{ \left[\frac{\partial}{\partial u} \frac{\partial L}{\partial x} \right]^T \delta x + \left[\frac{\partial}{\partial u} \frac{\partial L}{\partial u} \right] \delta u \right\}. \quad (4.14)$$

Введем более компактную запись в виде

$$\delta^2 L = \frac{1}{2} \delta x^T \delta u^T \begin{bmatrix} \frac{\partial}{\partial x} \frac{\partial L}{\partial x} & \frac{\partial}{\partial u} \frac{\partial L}{\partial x} \\ \left[\frac{\partial}{\partial u} \frac{\partial L}{\partial x} \right]^T & \frac{\partial}{\partial u} \frac{\partial L}{\partial u} \end{bmatrix} \begin{bmatrix} \delta x \\ \delta u \end{bmatrix}. \quad (4.15)$$

Если определить $\delta z' = [\delta x' \delta u']$,

$$P = \begin{bmatrix} \frac{\partial}{\partial x} \frac{\partial L}{\partial x} & \frac{\partial}{\partial u} \frac{\partial L}{\partial x} \\ \left[\frac{\partial}{\partial u} \frac{\partial L}{\partial x} \right]^T & \frac{\partial}{\partial u} \frac{\partial L}{\partial u} \end{bmatrix},$$

то выражение (4.15) сводится к

$$\delta^2 L = \frac{1}{2} \delta z^T P \delta z = \frac{1}{2} \|\delta z\|_P^2,$$

то есть к стандартной квадратичной форме. Положительно определенная квадратичная форма определяется как форма, для которой $\delta z^T P \delta z > 0$ для всех ненулевых δz . Положительно полуопределенная матрица P определяется из условия $\delta z^T P \delta z \geq 0$ для всех ненулевых δz . Точно так же определяются отрицательно определенные и отрицательно полуопределенные квадратичные формы и матрицы.

Таким образом, можно сформулировать два необходимых условия [15] наличия экстремума $J(x, u)$ в данном интервале X для выпуклой или вогнутой $J(x, u)$. Если $J(x, u)$ не является ни выпуклой, ни вогнутой, второе условие будет лишь достаточным.

Итак, имеем следующие условия экстремума:

1. Следующие векторы будут нулевыми

$$\frac{\partial L}{\partial x} = 0, \frac{\partial L}{\partial u} = 0.$$

2. Следующая матрица

$$\begin{bmatrix} \frac{\partial}{\partial x} \frac{\partial L}{\partial x} & \frac{\partial}{\partial u} \left(-\frac{\partial L}{\partial x} \right) \\ \left[\frac{\partial}{\partial u} \frac{\partial L}{\partial x} \right]^T & \frac{\partial}{\partial u} \frac{\partial L}{\partial u} \end{bmatrix}$$

определяется $\begin{cases} \text{положительно полуопределенной для минимума} \\ \text{при } f(x, u) = 0 \\ \text{отрицательно полуопределенной для максимума} \\ \text{при } f(x, u) = 0. \end{cases}$

Достаточное условие наличия минимума (максимума) функции, если первая вариация обращается в нуль, - это положительная (отрицательная) вторая вариация при нулевой первой [15]. Эти условия являются общими, и их следует модифицировать, если возможно оптимальное решение.

Например, имеется линейная система, описываемая уравнением

$$f(x, u) = Ax + Bu + C = 0,$$

требуется найти m -мерный вектор u , минимизирующий величину

$$J(x, u) = \frac{1}{2} \|u\|_R^2 + \frac{1}{2} \|x\|_Q^2,$$

где A – матрица размера $n \times n$;
 B – матрица размера $n \times m$;
 x , C и O – n – мерные векторы;
 R и Q – положительно определенные симметричные матрицы размера $n \times n$ и $m \times m$.

Функция Лагранжа формулируется следующим образом:

$$L = \frac{1}{2} U^T R u + \frac{1}{2} x^T Q x + \lambda^T [Ax + Bu + C].$$

Для минимизации J необходимо, чтобы

$$\frac{\partial L}{\partial x} = Qx + A^T \lambda = 0, \quad \frac{\partial L}{\partial u} = Ru + B^T \lambda = 0,$$

где λ выбирается таким образом, чтобы выполнялось заданное ограничение $Ax + Bu + C = O$.

Таким образом, находим, что

$$U = -(R + B^T A^{-T} Q A^{-1} B)^{-1} B^T A^{-T} Q A^{-1} C = -R^{-1} B^T (A Q^{-1} A^T + B R^{-1} B^T)^{-1} C$$

есть **оптимальный вектор u** .

Для определения того, действительно ли данное решение минимизирует $J(x, u)$, найдем вторую вариацию и проверим выполнение необходимого условия 2, изложенного выше. Из уравнения (4.14) и условия задачи получим

$$\delta^2 J = \frac{1}{2} [\delta x^T \delta u^T] \begin{bmatrix} Q & O \\ O & R \end{bmatrix} \begin{bmatrix} \delta x \\ \delta u \end{bmatrix} = \frac{1}{2} \delta x^T Q \delta x + \frac{1}{2} \delta u^T R \delta u.$$

Для того чтобы $J(x, u)$ имела минимум $\delta^2 J \geq 0$, достаточно, чтобы Q и R были неотрицательными определенными. Следующее условие получим, отметив, что первая вариация $f(x, u) = 0$ дает $A \delta x + B \delta u = 0$ и поэтому для $\delta^2 J > 0$ необходимо только, чтобы $R + B^T A^{-T} Q A^{-1} B$ было положительно определенным.

Далее рассмотрим на примере, как минимизировать функцию стоимости

$$J = \frac{1}{2} \|x\|_Q^2$$

при условии:

$$x + bu + c = 0,$$

где на скалярную величину управления наложено ограничение $|u| \leq 1$. Эту задачу можно решить, если ограничения на величину управления отсутствуют, следующим образом

$$u = -(b^T Q b)^{-1} b^T Q c.$$

Если $|u| < 1$, будем иметь то, что называют оптимальным (сингулярным) решением. Это происходит, потому, что функция L в переменной управления является линейной и $\partial L / \partial u = \lambda^T b = 0$ представляет собой уравнение для стационарной точки, которая вполне может быть точкой минимума. Если $b^T Q b$ положительно определена, она является по крайней мере локальным минимумом. Если полученная величина u лежит в указанных границах, то это значение и

есть решение задачи. Если значение по модулю больше 1, истинное решение для u должно быть на границе. Этот тип задач относится к теории **оптимального** управления.

Предположим, что исследуется постоянный вектор, искаженный шумом. Это можно записать следующим выражением

$$Z = Hx + v,$$

где Z - состоящий из элементов, есть m - мерный вектор;

H - матрица размера $m \times n$;

x - n - мерный вектор;

v - m - мерный вектор, представляющий собой шум измерения. Требуется получить наилучшую (оптимальную) оценку x , при которой величина

$$J = \frac{1}{2} \|z - H \dot{x}\|_{R^{-1}}^2 \text{ минимальна.}$$

Здесь R - симметричная положительно определенная матрица. Для этого приравняем

$$\frac{\partial J}{\partial \dot{x}} = H^T R^{-1} (z - H \dot{x}) = 0$$

Таким образом, **оптимальная** по критерию минимальной квадратичной ошибки оценка равна

$$\dot{x} = (H^T R^{-1} H)^{-1} H^T R^{-1} z.$$

Один из простейших случаев, представляющих интерес, возникает, когда находятся m скалярных оценок. При этом целесообразно представить H в виде единого m - мерного вектора, или, в виде вектора - столбца из единиц, а матрицу R в виде тождественной матрицы. Для этого простейшего случая для наилучшего оценивания x имеем

$$\dot{x} = \frac{H^T z}{m} = \frac{1}{m} \sum_{i=1}^m z_i, \quad (4.16)$$

что является хорошо известным выражением среднего числа наблюдений.

Представляет интерес случай, когда после вычисления оценки $\dot{x}$ появляется дополнительное измерение. Если H - матрица, например, порядка 1000×20 , то обращение и умножение матриц связано с трудоемкими операциями. Повторять эти операции для новой матрицы 1001×20 будет невозможно даже на ЭВМ, особенно если требуется расчет в «реальном» масштабе времени.

Таким образом, необходимо найти решение, позволяющее добавлять новое измерение, без повторения всего расчета полностью. Метод, используемый для этого, носит название **рекуррентного**, или метода **последовательного оценивания**. Эти методы играют важную роль в современной теории систем.

Пусть имеется совокупность измерений

$$z = Hx + v,$$

$$z = \begin{bmatrix} Z_1 \\ Z_2 \\ \dots \\ Z_m \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & \dots & h_{1n} \\ h_{21} & h_{22} & \dots & h_{2n} \\ \dots & \dots & \dots & \dots \\ h_{m1} & h_{m2} & \dots & h_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{bmatrix} + \begin{bmatrix} v_1 \\ v_2 \\ \dots \\ v_m \end{bmatrix},$$

причем $\dot{x}_m = (H^T R^{-1} H)^{-1} H^T R^{-1} z$. Предположим теперь, что получено дополнительное измерение, так что имеем

$$\begin{bmatrix} z \\ z_{m+1} \end{bmatrix} = \begin{bmatrix} H \\ h' \end{bmatrix} \begin{bmatrix} \dot{x}_m + \Delta x \end{bmatrix} + \begin{bmatrix} V \\ v_{m+1} \end{bmatrix}.$$

Задача заключается, в получении **оптимальной** (наилучшей) оценки x , $\dot{x}_{m+1}$, такой, что

$$J = \frac{1}{2} \left\| \begin{bmatrix} z \\ z_{m+1} \end{bmatrix} - \begin{bmatrix} H \\ h' \end{bmatrix} \dot{x}_{m+1} \right\|^2 \text{ минимально.}$$

Выполнив процедуру аналогичную предыдущей, получим **оптимальную** оценку x :

$$\dot{x}_{m+1} = \left(\begin{bmatrix} H \\ h' \end{bmatrix} \begin{bmatrix} H^T \\ h'^T \end{bmatrix} \right)^{-1} \begin{bmatrix} H \\ h' \end{bmatrix}^T \begin{bmatrix} z \\ z_{m+1} \end{bmatrix}$$

при упрощающем предположении, что матрица R является тождественной матрицей. Это означает присваивание всем изменениям равных весов. Используя лемму об обращении матиц, сформированную и доказанную в [18], можно составить рекурсивную процедуру. Заметим, что

$$\left\{ \begin{bmatrix} H \\ h' \end{bmatrix}^T \begin{bmatrix} H \\ h' \end{bmatrix} \right\} = [H^T H + h h^T]^{-1}. \quad (4.17)$$

Если определить

$$P_{m+1}^{-1} = \begin{bmatrix} H \\ h' \end{bmatrix}^T \begin{bmatrix} H \\ h' \end{bmatrix} = P_m^{-1} + h h^T,$$

то лемма об обращении матриц имеет вид

$$P_{m+1} = P_m - P_m h [h^T P_m h + 1]^{-1} h^T P_m, \quad (4.18)$$

что приводит к получению рекуррентной формулы:

$$\begin{aligned} \dot{x}_{m+1} &= P_{m+1} [H^T z + h z_{m+1}] = P_m H^T z + P_m h z_{m+1} - P_m h [h^T P_m h + 1]^{-1} h^T P_m [H^T z + h z_{m+1}] = \\ &= \dot{x}_m + P_m h [h^T P_m h + 1]^{-1} \left[z_{m+1} - h^T \dot{x}_m \right]. \end{aligned} \quad (4.19)$$

Таким образом, новая оценка равна старой плюс линейный поправочный коэффициент, учитывающий только новые данные и старое значение P_m . Для оценки скаляра x при условии, что H представляет собой единичный m -мерный вектор, имеем [1]

$$\begin{aligned} P_m^{-1} &= m, \quad P_{m+1} = \frac{1}{m+1}, \quad X_m = \frac{1}{m} \sum_{i=1}^m z_i, \\ \dot{x}_{m+1} &= \dot{x}_m + \frac{1}{m+1} \left[z_{m+1} - \dot{x}_m \right] = \dot{x}_m \left[\frac{m}{m+1} \right] + \frac{z_{m+1}}{m+1}, \end{aligned}$$

что и следовало ожидать в этом случае.

4.3. Динамическая оптимизация без ограничений

Рассмотрим задачу выбора из множества всех действительных дифференцированных функций на интервале $[t_0, t_f]$ функции $x: [t_0, t_f] \rightarrow R$, минимизирующей функцию стоимости

$$J(x) = \int_{t_0}^{t_f} \varphi[x(t), \dot{x}(t), t] dt. \quad (4.20)$$

Такие функции будем называть **допустимыми траекториями**. При этом всюду будем считать φ непрерывной в x , $\dot{x}$ и t и имеющей непрерывные частные производные по x и $\dot{x}$.

Следует отметить, что функция стоимости, представленная уравнением (4.20) и называемая формой Лагранжа, эквивалентна, при некоторых предположениях об однородности, ряду других полезных представлений функции стоимости, два из которых – это форма Больца

$$J'(x) = {}^0F[x(t_f), t_f] + \int_{t_0}^{t_f} A[x(t), \dot{x}(t), t] dt$$

и форма Майера $J''(x) = \sigma[x(t_f), t_f]$.

Форма Майера, которую можно записать так, чтобы σ также зависела от $x(t_0)$, представляет собой теоретический интерес благодаря своей простоте. Обе формы Майера и Лагранжа являются специальными случаями формы Больца. Видно, что при $\varphi = {}^0\dot{F} + A$ форма Больца переходит в форму Лагранжа при фиксированном ${}^0F[x(t_0), t_0]$. Чтобы показать, что форму Больца можно также записать в форме Майера, следует дополнить пространство состояний в задаче Больца переменной $\dot{x}^0(t) = A[x(t), x(t), t]$, где $x^0(t_0) = 0$, и определить $\sigma[x(t_f), t_f] = {}^0F[x(t_f), t_f] + x_0(t)$.

Далее перейдем к определению необходимых условий в форме уравнения Эйлера – Лагранжа и условий трансверсальности, при которых допустимая траектория должна быть оптимальной. Пусть $\dot{x}$ есть допустимая оптимальная траектория. Рассмотрим допустимую, но необязательно оптимальную траекторию $\bar{x}$, где для всех $t \in [t_0, t_f]$

$$x(t) = \dot{x} + \varepsilon \eta(t), \quad (4.21)$$

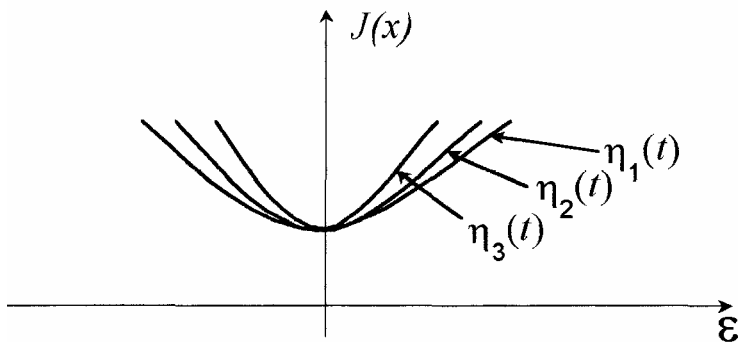


Рис. 4.4. Задача минимизации в вариационном исчислении

а $\eta(t)$ – вариация $x(t)$. ε – малая величина. Графики зависимости $J\langle x \rangle$ от ε для различных $\eta(t)$ показаны на рис. 4.4. Очевидно, что все кривые имеют минимум при $\varepsilon=0$, так как $\dot{x}(t) = x(t)|_{\varepsilon=0}$. Таким образом, на экстремумах имеем, что

$$\left. \frac{\partial J\langle x \rangle}{\partial \varepsilon} \right|_{\varepsilon=0} = 0 \quad (4.22)$$

не зависит от выбранного значения $\eta(t)$. Решение полученное из уравнения (4.21), может быть для $J\langle x \rangle$ точкой минимума, максимума или стационарной

точкой. Условие минимума – положительное значение $\frac{\partial^2 J}{\partial \mathcal{E}^2}$ при $\mathcal{E} = 0$ независимо от $\eta(t)$. Однако для большинства физических задач очевидно, что если уравнение (4.22) имеет решение, то это будет решение, минимизирующее (максимизирующее) интеграл $J\langle x \rangle$. Теперь можно найти экстремум выражения (4.20), используя уравнения (4.21) и (4.22). Дифференцируя уравнение (4.21) по t , получим

$$\overset{\vee}{x}(t) = \dot{x}(t) + \mathcal{E} \overset{\vee}{\eta}(t). \quad (4.23)$$

Если подставить уравнения (4.22) и (4.23) в заданный функционал (4.20), то получим

$$J\langle x \rangle = \int_{t_0}^{t_1} \varphi \left[\overset{\vee}{x}(t) + \mathcal{E} \eta(t), \overset{\vee}{x}(t) + \mathcal{E} \overset{\vee}{\eta}, t \right] dt \quad (4.24)$$

следует отметить, что

$$\lim_{\mathcal{E}=0} J(x) = J\left(\dot{x}\right), \quad \lim_{\mathcal{E}=0} x(t) = \dot{x}(t).$$

Поэтому для нахождения экстремалей $J(x)$ воспользуемся уравнением (4.22). Если $u = f(x, y, z \dots)$ является функцией нескольких переменных, каждая из которых представляет собой дифференцируемую функцию от $r, v, w, \dots$, то u как функция от этих новых независимых переменных является дифференцируемой и, следовательно, применимо следующее цепное правило:

$$\begin{aligned} \frac{\partial u}{\partial r} &= \frac{\partial u}{\partial x} \frac{\partial x}{\partial r} + \frac{\partial u}{\partial y} \frac{\partial y}{\partial r} + \dots \\ \frac{\partial u}{\partial v} &= \frac{\partial u}{\partial x} \frac{\partial x}{\partial v} + \frac{\partial u}{\partial y} \frac{\partial y}{\partial v} + \dots \end{aligned}$$

$$\left. \frac{\partial J\langle x \rangle}{\partial \mathcal{E}} \right|_{\mathcal{E}=0} = \int_{t_0}^{t_1} \left\{ \eta(t) \frac{\partial \varphi(\dot{x}, \overset{\vee}{x}, t)}{\partial \dot{x}} + \eta(t) \frac{\partial \varphi(\dot{x}, \overset{\vee}{x}, t)}{\partial \overset{\vee}{x}} \right\} dt = 0 \quad (4.25)$$

или

$$0 = \int_{t_0}^{t_1} \eta(t) \frac{\partial \varphi(\dot{x}, \overset{\vee}{x}, t)}{\partial \dot{x}} dt + \int_{t_0}^{t_1} \eta(t) \frac{\partial \varphi(\dot{x}, \overset{\vee}{x}, t)}{\partial \overset{\vee}{x}} dt. \quad (4.26)$$

После упрощения уравнение (4.26) примет вид

$$0 = \int_{t_0}^{t_f} \eta(t) \left[\frac{\partial \varphi}{\partial \dot{x}} - \frac{\partial}{\partial t} \frac{\partial \varphi}{\partial \dot{x}} \right] dt + \left. \frac{\partial \varphi}{\partial \dot{x}} \eta(t) \right|_{t_0}^{t_f} \quad (4.27)$$

[То есть используем формулу интегрирования по частям

$$\int_a^b u dv = uv \Big|_a^b - \int_a^b v du$$

и положив, что $u = \frac{\partial \varphi}{\partial \dot{x}}$, $dv = \dot{\eta}(t) dt$, $du = \frac{d}{dt} \frac{\partial \varphi}{\partial \dot{x}} dt$, $v = \eta(t)$,

получим
$$\int_{t_0}^{t_f} \eta(t) \frac{\partial \varphi}{\partial \dot{x}} dt = \left. \eta(t) \frac{\partial \varphi}{\partial \dot{x}} \right|_{t_0}^{t_f} - \int_{t_0}^{t_f} \eta(t) \frac{d}{dt} \frac{\partial \varphi}{\partial \dot{x}} dt]$$

Ясно, что если

$$\frac{\partial \varphi}{\partial \dot{x}} - \frac{d}{dt} \frac{\partial \varphi}{\partial \dot{x}} = 0, \quad (4.28)$$

$$\left. \frac{\partial \varphi}{\partial \dot{x}} \eta(t) \right|_{t_0}^{t_f} = 0 \text{ для } t = t_0, t_f \quad (4.29)$$

по **оптимальной траектории**, то уравнение (4.27) справедливо. Уравнение (4.27) должно выполняться для всех допустимых вариаций. Обратное также справедливо, то есть если уравнение (4.27) справедливо независимо от η , то уравнение (4.28) и (4.29) справедливы. Для случая, когда $x(t_0)$ и $x(t_f)$ фиксированы и, следовательно, $\eta(t_0) = \eta(t_f) = 0$, справедливость уравнения (4.21) доказывается с помощью следующей леммы.

Если $x(t)$ непрерывная на замкнутом интервале $t \in [t_1, t_2]$ и если $\int_{t_1}^{t_2} x(t) \eta(t) dt = 0$ для каждого $\eta(t)$, содержащегося в $[t_1, t_2]$, так что $\eta(t_1) = \eta(t_2) = 0$, то $x(t) = 0$ для всех t в $[t_1, t_2]$ [18].

Уравнение (4.28) и (4.29) – два очень важных соотношения, которые лежат в основе решения вариационных задач. Уравнение (4.28) известно как уравнение Эйлера – Лагранжа, а уравнение (4.29) – как связанное с ним условие трансверсальности. Эти уравнения определяют двухточечную краевую задачу, решение которой представляет собой допустимую **оптимальную траекторию**,

выраженную через известное φ . Здесь допускается, что производная $\frac{d}{dt} \frac{\partial \varphi}{\partial \dot{x}}$ непрерывна вдоль оптимальной траектории. Другой вывод следует в котором это допущение отсутствует [22]. Уравнение (4.21) можно так же представить в интегральном виде [13] $\frac{\partial \varphi}{\partial \dot{x}} \Big|_{t_0}^{t_f} \frac{\partial \varphi}{\partial x} dt + C$,

где C – произвольная постоянная.

Таким образом, для сформулированной задачи **оптимального управления** с фиксированным временем окончания, если $\dot{x}$ – допустимая оптимальная траектория, уравнение Эйлера – Лагранжа (4.28) и условия трансверсальности (4.29) должны быть справедливы для $\dot{x}$. Следует указать, что уравнения (4.28) и (4.29) являются лишь необходимыми условиями, поэтому траектория удовлетворяющая этим условиям не обязательно оптимальна.

4.4. Динамическая оптимизация с ограничениями в форме равенств

Рассмотрим следующую задачу с ограничениями: минимизировать функцию стоимости

$$J = \int_{t_0}^{t_f} \varphi(x, \dot{x}, t) dt \quad (4.30)$$

при наличии ограничения в форме m – мерного векторного равенства

$$A(x, \dot{x}, t) = 0 \text{ для всех } t \in [t_0, t_f]. \quad (4.31)$$

Траектория состояний считается допустимой, если кроме выполнения условий гладкости, сформулированных выше, выполняется равенство (4.31).

Можно показать, что эта задача с ограничениями эквивалентна задаче минимизации функции стоимости

$$J' = \int_{t_0}^{t_f} \left[\varphi(x, \dot{x}, t) + \lambda^T(t) A(x, \dot{x}, t) \right] dt \quad (4.32)$$

без ограничений, где меняющийся во времени m – мерный вектор $\lambda(t)$ есть вектор, эквивалентный множителю Лагранжа.

Для того чтобы проиллюстрировать применения метода множителей Лагранжа, рассмотрим случай, когда $\mathbf{X}$ представляет собой двумерный вектор. Пусть требуется минимизировать функционал (с фиксированными граничными точками)

$$J = \int_{t_0}^{t_f} \varphi(x_1, x_2, \overset{\vee}{x}_1, \overset{\vee}{x}_2) dt \quad (4.33)$$

при ограничении

$$A(x_1, x_2, t) = 0. \quad (4.34)$$

Необходимое условие минимума определяется равенством нулю первой вариации от (4.33)

$$\delta J = \int_{t_0}^{t_f} \delta x_1 \left\{ \left[\frac{\partial \varphi}{\partial x_1} - \frac{d}{dt} \frac{\partial \varphi}{\partial \overset{\vee}{x}_1} \right] + \delta x_2 \left[\frac{\partial \varphi}{\partial x_2} - \frac{d}{dt} \frac{\partial \varphi}{\partial \overset{\vee}{x}_2} \right] \right\} dt = 0. \quad (4.35)$$

Если δx_1 не зависит от δx_2 , можно просто приравнять каждый из двух членов в (4.35) нулю. Поскольку наличие ограничения подразумевает зависимость между x_1 и x_2 , следует принять во внимание заданное ограничение. Определив вариацию (4.34), получим

$$\delta A = \frac{\partial A}{\partial x_1} \delta x_1 + \frac{\partial A}{\partial x_2} \delta x_2 = 0. \quad (4.36)$$

Из этого также следует, что для любого $\lambda(t)$ можно умножить (4.36) на $\lambda(t)$ и проинтегрировать полученные произведения (выражения), так что

$$\int_{t_0}^{t_f} \lambda(t) \left[\frac{\partial A}{\partial x_1} \delta x_1 + \frac{\partial A}{\partial x_2} \delta x_2 \right] dt = 0 \quad (4.37)$$

Сложим (4.35) и (4.37), получим

$$0 = \int_{t_0}^{t_f} \left\{ \delta x_1 \left[\frac{\partial \varphi}{\partial x_1} - \frac{d}{dt} \frac{\partial \varphi}{\partial \overset{\vee}{x}_1} + \lambda \frac{\partial A}{\partial x_1} \right] + \delta x_2 \left[\frac{\partial \varphi}{\partial x_2} - \frac{d}{dt} \frac{\partial \varphi}{\partial \overset{\vee}{x}_2} + \lambda \frac{\partial A}{\partial x_2} \right] \right\} dt. \quad (4.38)$$

Выберем теперь такое значение λ , чтобы подынтегральное выражение в первых скобках было равно нулю. Из этого следует, что поскольку δx_2 произвольно, подынтегральное выражение во вторых скобках должно быть равно нулю. Очевидно, что можно было бы получить те же результаты если бы сформулировать данную задачу с добавлением к функции стоимости ограничения с помощью множителя Лагранжа, как в уравнении (4.32) и применить к этой функции стоимости уравнение Эйлера-Лагранжа. Результирующие уравнения Эйлера-Лагранжа должны быть затем решены с учетом ограничения в форме равенства (4.34) и т.д.

4.5. Динамическая оптимизация в форме неравенств с ограничениями

В задачах, представляющих интерес с позиции управления, имеются различные ограничения в форме неравенств на вектор управления. (Например, максимальная реактивная тяга ограниченная максимальной входной реактивностью ядерного реактора). Наличие ограничений в форме неравенств необходимо учитывать при разработке оптимальных систем. Таким образом имеем дело с минимизацией функции стоимости вида

$$J = \int_{t_0}^{t_f} \varphi \left(x, \dot{x}, t \right) dt \quad (4.39)$$

с ограничениями в форме равенств

$$A \left(x, \dot{x}, t \right) = 0 \quad (4.40)$$

и в форме неравенств

$$\pi_{\min} \leq \pi \left(x, \dot{x}, t \right) \leq \pi_{\max}. \quad (4.41)$$

Для того чтобы быть допустимой, траектория должна удовлетворять уравнениям (4.40) и (4.41), а также условиям гладкости, сформулированными выше. Для задачи с ограничениями в форме неравенств имеется несколько методов, один из которых – метод замены переменных. При этом выполняется преобразование ограничения в форме неравенства (4.41) в ограничение в форме равенства путем введения новых переменным состояния γ_i , удовлетворяющих уравнениям

$$(\pi_{\max_i} - \pi_i)(\pi_i - \pi_{\min_i}) = \gamma_i^2, \quad i = 1, 2, 3, \dots, \quad (4.42)$$

легко показать, что уравнение (4.42) эквивалентно (4.41), так как для того чтобы существовало γ_i , $i = 1, 2$, нужно, чтобы было справедливо (4.41) и наоборот. Теперь, чтобы внести ограничения (4.40) и (4.42) в (4.39), можно использовать множители Лагранжа. Потом становится возможным использовать обычные необходимые условия экстремума.

Например, рассмотрим динамику системы:

$$\dot{x}_1 = x_2(t), \quad \dot{x}_2 = u(t)$$

при начальных условиях $x_1(t_0) = x_0$ и $x_2(t_0) = v_0$.

Задача состоит в том, чтобы найти такое уравнение, которое максимизирует $x_1(t_f)$ для фиксированного t_f при граничном условии $x_2(t_f) = v$, и ограничениях в форме неравенства в ограничение в форме равенства, введя новую пе-

ременную $\alpha(t)$ и заменив ограничения в форме неравенства на $(u - u_{\min})(u_{\max} - u) - \alpha^2 = 0$.

Таким образом, задача может быть представлена как задача минимизации $J = -x_1(t_f)$ при наличии ограничений в форме равенств

$$\begin{aligned} \dot{x}_1 &= x_2(t), & x_1(t_0) &= x_0, & x_1(t_f) &- \text{произвольно} \\ \dot{x}_2 &= u(t), & x_2(t_0) &= v_0, & x_2(t_f) &= v_f \\ (u - u_{\min})(u_{\max} - u) - \alpha^2 &= 0. \end{aligned}$$

Функция стоимости с множителями Лагранжа принимает вид

$$J' = -x_1(t_0) + \int_{t_0}^{t_f} \left\{ -\dot{x}_1 + \lambda_1 [x_2 - \dot{x}_1] + \lambda_2 [u - \dot{x}_2] + \lambda_3 [(u - u_{\min})(u_{\max} - u) - \alpha^2] \right\} dt.$$

Из уравнения Эйлера-Лагранжа

$$\frac{d}{dt} \frac{\partial \Phi}{\partial \dot{x}} - \frac{\partial \Phi}{\partial x} = 0, \quad x^T = [x_1, x_2, u]$$

при $\Phi = \lambda_1 [x_2 - \dot{x}_1] + \lambda_2 [u - \dot{x}_2] + \lambda_3 [(u - u_{\min})(u_{\max} - u) - \alpha^2] - \dot{x}_1$

находим

$$\begin{aligned} \dot{\lambda}_1 &= 0, & \dot{\lambda}_2 &= -\lambda_1, \\ O &= -\lambda_2 + \lambda_3 [2u - u_{\max} - u_{\min}], & O &= \alpha \lambda_3. \end{aligned}$$

Для определения единственного недостающего условия достижения для $x_1(t_f)$ применим естественное граничное условие (условие трансверсальности).

Имеем

$$\left. \frac{\partial \Phi}{\partial x_1} \right|_{t=t_f} = 0 = -1 - \lambda_1(t_f).$$

Таким образом, подошли к двухточечной краевой задаче, решение которой определяет **оптимальные переменные** состояния и управления, а именно:

$$\begin{aligned} \dot{x}_1 &= x_2(t), & x_1(t_0) &= x_0, \\ \dot{x}_2 &= x(t), & x_2(t_0) &= v_0 \\ \dot{\lambda}_1 &= 0, & \lambda_1[t_f] &= -1, \\ \dot{\lambda}_1 &= -\lambda_1(t), & x_2(t_f) &= v_f, \\ \alpha(t) \lambda_3(t) &= 0, \\ \lambda_2(t) &= \lambda_3(t) [2u(t) - u_{\max} - u_{\min}], \\ \alpha^2(t) &= [u(t) - u_{\min}] [u_{\max} - u(t)]. \end{aligned}$$

Данная задача является нелинейной из-за введения последних трех уравнений и получить ее решение достаточно сложно, если не использовать ЭВМ. В обычном варианте этой задачи $u_{\min} = -1$ и $u_{\max} = +1$. В этом случае можно показать, что $\alpha(t) = 0$ и $u(t) = -\text{sign}\lambda_2(t)$,

где

$$\begin{aligned} \text{sign}\lambda_2 &= 1, \text{ если } \lambda_2 > 0, \\ \text{sign}\lambda_2 &= -1 \text{ если } \lambda_2 < 0. \end{aligned}$$

Однако это не меняет нелинейного характера двухточечной краевой задачи.

4.6. Анализ задачи оптимального управления без ограничений в форме неравенств

Рассмотрим задачу определения допустимой функции управления u с целью минимизации функционала

$$J = {}^0F[x(t), t] \Big|_{t=t_0}^{t=t_f} + \int_{t_0}^{t_f} \Phi[x(t), u(t), t] dt, \quad (4.43)$$

причем 0F и Φ имеют непрерывные частные производные по x и u .

Для того, чтобы в функции стоимости учесть системное ограничение, заданное в виде дифференцированного уравнения, воспользуемся методом множителей Лагранжа

$$J = {}^0F[x(t), t] \Big|_{t=t_0}^{t=t_f} + \int_{t_0}^{t_f} \{ \Phi[x(t), u(t), t] + \lambda^T(t) [f[x(t), u(t), t] - \dot{x}] \} dt.$$

Определим скалярную функцию Гамильтона в виде

$$H[x(t), u(t), \lambda(t), t] = \Phi[x(t), u(t), t] + \lambda^T(t) f[x(t), u(t), t]. \quad (4.45)$$

Таким образом, функция стоимости принимает вид

$$J = {}^0F[x(t), t] \Big|_{t=t_0}^{t=t_f} + \int_{t_0}^{t_f} \{ H[x(t), u(t), \lambda(t), t] - \lambda^T(t) \dot{x} \} dt. \quad (4.46)$$

Проинтегрировав последний член подынтегрального выражения (4.46), получим

$$J = \{ {}^0F[x(t), t] - \lambda^T(t) x(t) \} \Big|_{t=t_0}^{t=t_f} + \int_{t_0}^{t_f} \{ H[x(t), u(t), \lambda(t), t] + \lambda^T(t) \dot{x}(t) \} dt. \quad (4.47)$$

Определим первую вариацию J

$$\delta J = \left\{ \delta x^T \left[\frac{\partial {}^0F}{\partial x} - \dot{\lambda} \right] \right\} \Big|_{t=t_0}^{t=t_f} + \int_{t_0}^{t_f} \left\{ \delta x^T \left[\frac{\partial H}{\partial x} + \dot{\lambda} \right] + \delta u^T \left[\frac{\partial H}{\partial u} \right] \right\} dt. \quad (4.48)$$

Необходимое условие минимума состоит в том, что для произвольных вариаций δx и δu первая вариация J должна быть равна нулю. Отсюда получаем существенное и необходимое условие существования минимума

$$\delta x' \left[\frac{\partial^0 F}{\partial x} - \lambda \right] = 0 \text{ для } t = t_0, t_f, \quad (4.49)$$

$$\dot{\lambda} = -\frac{\partial H}{\partial x}, \quad \dot{x} = f(x, u, t), \quad (4.50)$$

$$\frac{\partial H}{\partial u} = 0. \quad (4.51)$$

Рассмотрим подробнее условия трансверсальности, заданные выражением (4.49).

Для широкого класса задач оптимального управления начальное состояние системы определено, а состояние в момент достижения не определено. В этом случае из (4.49) получаем следующее условия трансверсальности:

$$x(t) = x_0, \quad \lambda(t_f) = \frac{\partial^0 F[x(t_f), t_f]}{\partial x(t_f)}, \quad (4.52)$$

поскольку $\delta x(t_0) = 0$, $x(t_0)$ фиксировано, а $\delta x(t_f)$ - произвольно. В другом широком классе задач и $x(t_0)$ и $x(t_f)$ фиксированы. В этом случае $\delta x(t_0)$ и $\delta x(t_f)$ должны равняться нулю, а $x(t_0)$ и $x(t_f)$ являются граничными условиями в двухточечной краевой задаче. Для многих задач оценивания ни $x(t_0)$, ни $x(t_f)$ не фиксированы, ${}^0F = 0$. В этом случае из (4.49) следует, что $\lambda(t_0) = \lambda(t_f) = 0$, это и будут граничные условия для данной задачи, так как $\delta x(t_0)$ и $\delta x(t_f)$ - произвольны. В другом случае можем иметь $x(t_0) = x_0$, ${}^0F = 0$ и $\|x(t_f)\|^2 = 1$. При этом легко показать, что окончательные условия трансверсальности получаются, если решить два скалярных уравнения, каждое из которых содержит n переменных:

$$\delta x^T(t_f)x(t_f) = 0, \quad \delta x^T(t_f)\lambda(t_f) = 0. \quad (4.53)$$

Дадим более точную и общую формулировку условий трансверсальности. В общем случае, когда в исходном положении

$$M[x(t_0), t_0] = 0, \quad (4.54)$$

а в конечном

$$N[x(t_f), t_f] = 0, \quad (4.55)$$

введем последние два условия в функцию 0F с помощью множителей Лагранжа ξ и v . Тогда функция стоимости приобретает следующий вид:

$$J = {}^0F[x(t), t]_{t=t_0}^{t=t_f} - \xi^T M[x(t_0), t_0] + v^T N[x(t_f), t_f] + \int_{t_0}^{t_f} \left\{ H[x(t), u(t), \lambda(t), t] - \lambda^T(t), \overset{\vee}{x} \right\} dt. \quad (4.56)$$

Для получения условий трансверсальности в начальный момент времени воспользуемся обычными вариационными методами:

$$\lambda(t_0) = \frac{\partial {}^0F}{\partial x} + \left(\frac{\partial M^T}{\partial x} \right) \xi, \quad M[x(t), t] = 0, \quad t = t_0. \quad (4.57)$$

Отсюда получаются n начальных условий с r параметрами, которые определяются из (4.57) так, чтобы выполнялись условия (4.54). Аналогично n условий в момент достижения получаются из

$$\lambda(t_f) = \frac{\partial {}^0F}{\partial x} + \frac{\partial N^T}{\partial x} v, \quad N[x(t), t] = 0, \quad t = t_f; \quad (4.58)$$

при q параметрах v так, чтобы выполнялось q условий (4.55).

Будем называть n – мерное векторное дифференциальное уравнение (4.50) сопряженным уравнением. Уравнение (4.51) дает связь между исходной динамикой системы

$$\overset{\vee}{x}(t) = f[x(t), u(t), t] \quad (4.59)$$

и сопряженным уравнением $\overset{\vee}{\lambda}$ (4.50). Это уравнение было получено из слагаемого

$$\delta J = \dots + \int_{t_0}^{t_f} \left\{ \delta u^T \frac{\partial H}{\partial u} + \dots \right\} dt.$$

Важно отметить, что δu должно быть произвольно, тогда можно для получения оптимального управления записать $\frac{\partial H}{\partial u} = 0$. Это условие выполняется для поставленной здесь задачи, в которой множество допустимых управлений δu бесконечно. При ограниченном множестве управляющих δu не может быть произвольным и условие $\frac{\partial H}{\partial u} = 0$ может оказаться некорректным. Решение представляет собой специальный случай принципа максимума Понтрягина.

Поскольку $H = \Phi + \lambda^T f$, то можно вычислить полную производную по времени

$$\frac{dH}{dt} = \frac{\partial \Phi}{\partial t} + \overset{\vee}{x}^T \left[\frac{\partial \Phi}{\partial x} + \left(\frac{\partial f^T}{\partial x} \right) \lambda \right] + \overset{\vee}{u}^T \left[\frac{\partial \Phi}{\partial u} + \left(\frac{\partial f^T}{\partial u} \right) \lambda \right] + \overset{\vee}{\lambda}^T f + \lambda^T \frac{\partial f}{\partial t}, \quad (4.60)$$

но из (4.50) и (4.45) имеем

$$\dot{\lambda} = -\frac{dH}{dx} = \frac{\partial \Phi}{\partial x} - \left(\frac{\partial f^T}{\partial u} \right) \lambda, \quad (4.61)$$

а из (4.50) и (4.45)

$$\frac{dH}{du} = \frac{\partial \Phi}{\partial u} + \left(\frac{\partial f^T}{\partial u} \right) \lambda. \quad (4.62)$$

Таким образом, поскольку $\dot{x} = \dot{\lambda} = \dot{\lambda}^T f$, равенство (4.60) переходит в форму

$$\frac{dH}{dt} = \frac{\partial \Phi}{\partial t} + \lambda^T \frac{\partial f}{\partial t} + u^T \frac{dH}{du}. \quad (4.63)$$

Видно, что если Φ и f не являются явными функциями времени Гамильтона, то они будут постоянными вдоль всей оптимальной траектории, где $\partial H / \partial u = 0$. Следует показать, что на оптимальной траектории это всегда справедливо, даже если не использовать условие $\partial H / \partial u = 0$.

Условие минимума J является неотрицательность второй вариации J вдоль всех траекторий так, чтобы было справедливо (4.59). Поэтому следует вычислить вторую вариацию J и приравнять нулю вариацию от уравнения (4.59)

$$\delta x - \left(\frac{\partial f}{\partial x} \right) \delta x - \left(\frac{\partial f}{\partial u} \right) \delta u = 0. \quad (4.64)$$

Используя это условие и ограничиваясь квадратичным членом разложения в ряд Тейлора $J(x + \delta x, u + \delta u) - J(x, u)$, получим для второй вариации выражение

$$\delta^2 J = \frac{1}{2} \left[\delta x^T \frac{\partial^2 F}{\partial x^2} \delta x \right]_{t=t_0}^{t=t_f} + \frac{1}{2} \int_{t_0}^{t_f} \left[\delta x^T \delta u^T \right] \begin{bmatrix} \frac{\partial^2 H}{\partial x^2} & \frac{\partial}{\partial u} \frac{\partial H}{\partial x} \\ \left[\frac{\partial}{\partial u} \frac{\partial H}{\partial x} \right]^T & \frac{\partial^2 H}{\partial u^2} \end{bmatrix} \begin{bmatrix} \delta x \\ \delta u \end{bmatrix} dt, \quad (4.65)$$

которое должно быть неотрицательным. Такое условие выполняется в том случае, если квадратная матрица порядка $n + m$ под знаком интеграла и матрица $\partial^2 F / \partial x^2$ неотрицательно определенные.

4.7. Анализ задачи непрерывного оптимального управления в фиксированные и неопределенные моменты достижения

Обобщим материал, изложенный выше для случая, когда время достижения не определено. Предположим, что начальный момент и вектор начального состояния определены. Далее можно получить решение для случая, когда упо-

мянутые величины не определены. Тогда задача сведется к минимизации функции стоимости

$$J = {}^0F[x(t_f), t_f] + \int_{t_0}^{t_f} \Phi[x(t), u(t), t] dt \quad (4.66)$$

для системы, описываемой уравнением

$$\dot{x} = f[x(t), u(t), t] \quad x(t_0) = x_0 \quad (4.67)$$

где t_0 фиксировано и в момент достижения $t = t_f$ справедливо q – мерное векторное уравнение

$$N[x(t_f), t_f] = 0. \quad (4.68)$$

Граничные условия $x(t_f) = Ct_f$, определенное выше, переходит в более общее $N[x(t_f), t_f] = 0$. Введем в функцию стоимости с помощью множителей Лагранжа ограничения, заданные в форме равенств

$$J = {}^0F[x(t_f), t_f] + v^T N[x(t_f), t_f] + \int_{t_0}^{t_f} \{ \Phi[x(t), u(t), t] + \lambda^T(t) f[x(t), u(t), t] - \dot{x}^T \} dt \quad (4.69)$$

Определим функцию Гамильтона

$H[x(t), u(t), \lambda(t), t] = \Phi[x(t), u(t), t] + \lambda^T(t) f[x(t), u(t), t]$ и преобразуем интеграл (4.69). В результате получим

$$J = {}^0F[x(t_f), t_f] + v^T N[x(t_f), t_f] - \lambda^T(t_f) x(t_f) + \lambda^T(t_0) x(t_0) + \int_{t_0}^{t_f} \left\{ H[x(t), u(t), \lambda(t), t] + \dot{\lambda}^T x(t) \right\} dt. \quad (4.70)$$

Первая вариация получается при

$$x(t) = \dot{x}(t) + h(t), \quad u(t) = \dot{u}(t) + \delta u(t), \quad t_f = \dot{t}_f + \delta t_f. \quad (4.71)$$

Далее определим разность $J[x, u, t_f] - J[\dot{x}, \dot{u}, \dot{t}_f]$ и будем учитывать только линейные члены разложения. Таким образом, опустим для упрощения значок пересечения ($\wedge$), получим

$$\begin{aligned} \delta J = & \delta t_f \left\{ H[x(t_f), u(t_f), \lambda(t_f), t_f] + \frac{\partial {}^0F'}{\partial t_f} \right\} + \delta x^T(t_f) \left\{ \frac{\partial {}^0F'}{\partial x} - \lambda(t_f) \right\} + \\ & + \int_{t_0}^{t_f} \left\{ h^T(t) \left[\frac{\partial H}{\partial x} + \dot{\lambda} \right] + \delta u^T(t) \left[\frac{\partial H}{\partial u} \right] \right\} dt, \end{aligned} \quad (4.72)$$

где

$${}^0F[x(t_f), v, t_f] = {}^0F[x(t_f), t_f] + v^T N[x(t_f), t_f]. \quad (4.73)$$

Для получения необходимого условия минимума следует приравнять нулю первую вариацию. Тогда получим систему $2n$ дифференцированных уравнений, определяющих оптимальное управление и вектор состояния

$$H = \Phi[x(t), u(t), t] + \lambda^T(t) f[x(t), u(t), t], \quad (4.74)$$

$$\frac{\partial H}{\partial \lambda} = \dot{x} = f[x(t), u(t), t], \quad (4.75)$$

$$\frac{\partial H}{\partial x} = -\dot{\lambda} = \frac{\partial f^T[x(t), u(t), t]}{\partial x} \lambda(t) + \frac{\partial \Phi[x(t), u(t), t]}{\partial x}, \quad (4.76)$$

$$\frac{\partial H}{\partial u} = 0 = \frac{\partial \Phi[x(t), u(t), t]}{\partial u} + \frac{\partial f^T[x(t), u(t), t]}{\partial u} \lambda(t). \quad (4.77)$$

Условия в начальный момент

$$x(t_0) = x_0, \quad (4.78)$$

а в конечный момент

$$\lambda(t_f) = \frac{\partial {}^0 F'}{\partial x(t_f)} = \frac{\partial {}^0 F}{\partial x(t_f)} + \left[\frac{\partial N^T}{\partial t_f} \right], \quad (4.79)$$

$$N \left[x(t_f), t_f \right] = 0 \quad (4.80)$$

и

$$H[x(t_f), u(t_f), \lambda(t_f), t_f] + \frac{\partial {}^0 F}{\partial t_f} + \left(\frac{\partial N^T}{\partial t_f} \right) v = 0. \quad (4.81)$$

Уравнение (4.79) дает n условий, которые входят q множителей Лагранжа, подлежащих определению. Уравнение (4.80) дает q уравнений для исключения множителей Лагранжа, а (4.81) дает еще одно дополнительное уравнение, которое необходимо для определения времени достижения.

Рассмотрим эти теоретические исследования на практических примерах:

4.7.1. Для интеграла, описываемого уравнением первого порядка $\dot{x} = u$, $x(0) = 1$, требуется найти **управление** $u(t)$, причем управление **оптимальное**, при котором $x(t_f) = 0$, где t_f не фиксировано, так чтобы при заданных α и β

$$J = t_f^\alpha = \frac{1}{2} \beta \int_0^{t_f} u^2 dt$$

было минимально. Для этой задачи

$$N[x(t_f), t_f] = x(t_f) = 0, \quad \Phi = \frac{1}{2} \beta u^2, \quad {}^0 F = t_f^\alpha, \quad H = \frac{1}{2} \beta u^2 + \lambda u.$$

Канонические уравнения $\dot{x} = u = -\lambda \beta, \quad \dot{\lambda} = 0$

с ограниченным условием $x(0) = 1$, $x(t_f) = 0$, где конечный момент времени определяется решением уравнения (4.81), которое в данном случае будет иметь вид

$$-\frac{\lambda^2(t_f)}{2\beta} + \alpha t_f^{\alpha-1} = 0.$$

Решение канонических уравнений:

$$x(t) = -\lambda(t_f) \frac{t}{\beta} + 1, \quad \lambda(t) = \lambda(t_f).$$

Но, поскольку $x(t_f) = 0$, $t_f = \beta \lambda^{-1}(t_f)$ и в частном случае, когда $\beta = \alpha = 1$, ясно, что $\lambda(t_f) = +\sqrt{2}$, это и будет решением для данного примера. Оптимальное управление, $u(t) = -\lambda(t) = -\sqrt{2}$, соответствующая траектория $x(t) = 1 - \sqrt{2} \cdot t$ при $t_f = 1/\sqrt{2}$.

4.7.2. Рассмотрим задачу о «минимальном времени». В этом случае

$${}^0F[x(t_f), t_f] = t_f, \quad \Phi = 0,$$

а **оптимальные управления** и соответствующая траектория определяются решениями уравнений (4.74) – (4.77), которые для рассматриваемой задачи имеют вид

$$H[x(t), u(t), \lambda(t), t] = \lambda^T(t) f[x(t), u(t), t],$$

$$\frac{\partial H}{\partial \lambda} = \dot{\mathbf{x}} = f[x(t), u(t), t],$$

$$\frac{\partial H}{\partial \mathbf{x}} = -\dot{\boldsymbol{\lambda}} = \frac{\partial f^T[x(t), u(t), t]}{\partial \mathbf{x}} \boldsymbol{\lambda}(t),$$

$$\frac{\partial H}{\partial u} = 0 = \frac{\partial f^T[x(t), u(t), t]}{\partial u} \boldsymbol{\lambda}(t)$$

с граничными условиями, определяемые по выражениям (4.78) – (4.81):

$$\mathbf{x}(t_0) = \mathbf{x}_0,$$

$$\lambda(t_f) = \frac{\partial N^T}{\partial x(t_f)} v,$$

$$N[x(t_f), t_f] = 0,$$

$$H[x(t_f), u(t_f), t_f] = -1 - \left(\frac{\partial N^T}{\partial t_f} \right) v.$$

Во многих случаях представляет интерес определение момента времени перехода системы в состояние покоя, так что

$$N[x(t_f), t_f] = x(t_f) = 0.$$

Тогда предыдущее выражение сведется к

$$x(t_0) = x_0, \quad x(t_f) = 0 \\ H[x(t_f), u(t_f), \lambda(t_f), t_f] = -1.$$

Если функция Гамильтона не является явной функцией времени, то из выражения (4.63), которое можно здесь применить, следует $\frac{dH}{dt} = 0$; поэтому для задачи о минимальном времени

$$N[x(t_f), u(t_f), \lambda(t_f), t_f] = H[x(t_f), u(t_f), \lambda(t_f)] = -1.$$

Таким образом, здесь не решается обычная задача о минимальном времени, так как на переменные управления (или состояния) не было наложено никаких ограничений в форме неравенств. Здесь можно рассматривать другой случай задачи, в котором ${}^0F=0$ и $\Phi=1$. При этом функция Гамильтона измениться, а оптимальные управления и вектор состояния, конечно, не меняются. Проведенные выше исследования **оптимального** решения (управления) по анализу систем дают основания полагать, что применение математических моделей процессов и систем (априорное их описание) значительно облегчают работу проектировщиков, разработчиков и системотехников при создании современных систем управления и принятии оптимальных управленческих решений при организации процессов и систем автоматизированного проектирования и управления разработками. Так как математическое моделирование процессов и систем является главной стадией и **функцией** (полезным действием) организации и создания современных процессов и систем автоматизации, то их **электронное моделирование** является **производной** от этой процедуры, что существенно сокращает затраты на доработку, корректировку и адаптацию различных процессов и систем автоматизации в промышленном производстве.

5. ОПТИМИЗАЦИЯ СОСТАВА ЭЛЕКТРОННОЙ СИСТЕМЫ УПРАВЛЕНИЯ ДВИГАТЕЛЯМИ ПО ИЗМЕНЕНИЯМ ПАРАМЕТРОВ

На основании проведенных проверок сравнением (верификации) основных методов анализа оптимального управления процессами и системами, используем эти методы на примере гражданских систем, а именно, оптимизируем состав электронной системы управления двигателями автомобилей. То есть, любую систему будь то: автомобиль, трактор, комбайн, электровоз, тепловоз или электронная система управления двигателями и др., следует отнести к **техническим системам**, которые обладают двумя свойствами: **затратностью и доходами**, то есть при исследовании такой технической системы (разработки вновь или модернизации существующих систем) следует, что **затрачивая средства** на разработку или модернизацию такой технической системы преследуется одна цель — получение **дохода**, а желательно **максимального дохода**. Необходимо отметить, что в каждом конкретном случае этот процесс представляет собой задачу, сложную как в практическом, так и в теоретическом отношении. Но если такая задача поставлена и решена, то есть выявлены и обоснованы значимые параметры указанных составляющих системы, то начинается следующий этап - установление и описание связей между затратами и доходом на ее обслуживание. Здесь наряду с другими могут быть использованы экономические - математические методы, либо методы функционально-стоимостной инженерии. Применение их предполагает разумный уровень априорного знания о системе исследования. Так например, использование критериев **«функциональность»** и **«стоимость»** требует определить общий характер функционирования системы, её назначение и априорную **полезность**. Теория полезности [13] требует введения того или иного отношения предпочтения на множестве параметров системы.

Это приводит к появлению некоторой **критической функции - оптимальной полезности**, которая позволяет изменять **полезность** и выбирать из допустимого множества решений функции, оптимальный набор априорных параметров.

Выбор наиболее подходящего и объективного метода (или способа) описания функции управления двигателями, критериев **функциональности и стоимости**, мотивирующих этот выбор является одной из важных задач для оптимизации технических и качественных решений ЭСУД. Причём такие критерии, как **функциональность и стоимость**, могут уточнить или расширить область применения методов оптимизации, найти новые и полезные интерпретации для сугубо гражданских объектов, оптимизации их состава и описания таких, как электронная система управления двигателями автомобилей любых модификаций.

5.1 Экономическо-математический метод оптимизации решений состава ЭСУД на основе законов классической математики

Общая формулировка, например, математического программирования [5] следующая: пусть требуется найти решение (в общем случае конечное количество функций от времени) $\bar{U}(t)$, при котором минимизируется (или максимируется) критерий, представляющий собой функцию от решения $\bar{U}(t)$, фазовых координат $\bar{x}$ и времени:

$$\min_{\bar{U}(t)} K = \min \Phi[\bar{U}(t), t, \bar{X}]. \quad (5.1)$$

Фазовые координаты – набор переменных, которые полностью определяют состояние системы. Например, если речь идёт об ЭСУД, то это комплексные характеристики выполнения плана выпуска машин с ЭСУД, запасных частей (подсистем), оборудования для ремонта и обслуживания, и т.д.

При этом должны удовлетворяться уравнения связей (т.е. модели процесса):

$$\partial[\bar{X}, \bar{X}, \dots, \bar{X}^{(m)}, t, \bar{u}(t)] = 0, \quad (5.2)$$

а также начальные и конечные условия:

$$\partial[\bar{X}_0, \bar{X}_0, \dots, \bar{X}_0^{(m)}, t_0, \bar{u}(t_0)] = 0. \quad (5.3)$$

Кроме того, должны соблюдаться ограничения по возможным значениям фазовых координат и решений:

$$\bar{X} \in \bar{X}_{oon.}, \quad \bar{U} \in \bar{U}_{oon.}, \quad (5.4)$$

(то есть вектор фазовых координат должен лежать в области допустимых значений).

В зависимости от тех или иных упрощений этой общей постановки могут применяться те или иные методы математического программирования.

Дифференциальное исчисление используется, если:

1. Критерий выражается в форме аналитической функции от решения (обслуживания):

$$K = \Phi(u_1, u_2, \dots, u_n), \quad (5.5)$$

то есть управление связями, начальные и граничные условия удаётся разрешить относительно фазовых координат, после чего исключить их из уравнения (5.1);

2. ограничения на фазовые координаты и решения отсутствуют;
3. решение представляет собой набор величин (но не функций от времени), то:
4. оптимальное решение может быть найдено из системы уравнений представляющих собой необходимые условия оптимальности:

$$\begin{aligned} \partial \Phi(u_1, u_2, \dots, u_n) &= 0; \\ \frac{\partial \Phi(u_1, u_2, \dots, u_n)}{\partial u_n} &= 0. \end{aligned} \quad (5.6)$$

Решив систему уравнений, и находят значения $u_1, u_2, \dots, \overline{u_n}$, удовлетворяющие необходимым условиям.

Но существуют ещё условия достаточности, анализ которых позволяет установить, имеется ли в найденной точке максимум, минимум или нет ни того ни другого.

Не останавливаясь на описании этих условий, отметим, что их проверку можно заменить численными расчётами: нужно несколько изменить найденное из необходимых условий u_i и посмотреть, как изменится K .

Линейное программирование может быть применено в тех случаях, когда критерий является линейной функцией решений (уравнений):

$$K = \sum_{k=1}^{k_1} C_k U_k \quad (5.7)$$

и имеются только линейные ограничения:

$$\sum_{k=1}^{k_1} a_{ik} U_k \leq b_i \quad \text{для } i=1, 2, \dots, m, \quad (5.8)$$

$$U_k \geq 0 \quad \text{для } k=1, 2, \dots, K_1 \quad (5.9)$$

Для решения этих уравнений разработано много способов, в том числе на ПЭВМ. Отметим, что этот метод – частный случай и принципа максимума, и динамического программирования. Специфика его – учёт большого числа ограничений.

Задача выбора оптимального типажа формулируется следующим образом. Пусть задана дифференциальная $\varphi(x)$ или интегральная

$F(x) = \int_0^x \varphi(x) dx$ функция потребности в чём-либо (в электронных системах

управления двигателями автомобилей, потребность в дополнительной информации об оборудовании, инструменте или, например, на сервисном обслуживании – потребность в каком-либо оборудовании). Эти функции могут быть получены обработкой статистических данных и их прогнозированием. Их вид изображён на рис.5.1. Пусть далее задана стоимость разработки $C_p(x)$, производства (или обслуживания) $C_o(x)$ и эксплуатации $C_e(x)$, рассматриваемого (или оборудования для ремонтного предприятия, или информации для системы управления двигателями) как функция от аргумента x .

Вид $C_o(x)$ показан на рис. 5.1.

Требуется определить оптимальное количество типов (или оборудования, или количество запасных частей для электронных системах управления двигателями) N , оптимальные значения аргумента для, каждого типа - набор x_k b_m $k=1$ до $k=N$ и количество (чего-то) каждого типа, для которых минимизируются суммарные затраты.

Математическая формулировка задачи следующая. Найти набор x_k , включая и их число минимизирующие суммарные затраты:

$$\begin{aligned} \min_{x_1, x_2, \dots, x_N} C = & \sum_{k=0}^N [F(x_{k+1}) - F(x_k)] C_o(x_{k+1}) + \sum_{k=0}^N C_p(x_{k+1}) + \\ & + \int_0^T \sum_{k=0}^N C_s(x_{k+1}) [F(x_{k+1}) - F(x_k)] dt. \end{aligned} \quad (5.10)$$

Несмотря на внешне простую запись, задача оказывается достаточно сложной.

Рассмотрим упрощённый случай одномерной задачи, когда удаётся получить аналитическое решение. Пусть:

$$F(x) = b[2x_N x - x^2] \text{ при } 0 \leq x \leq x_N; \quad (5.11)$$

$$F(x) = 0 \text{ при } x \leq 0; \quad (5.12)$$

$$F(x) = bx_N^2 \text{ при } x \geq x_N; \quad (5.13)$$

Именно такой случай показан на рис.5.1.

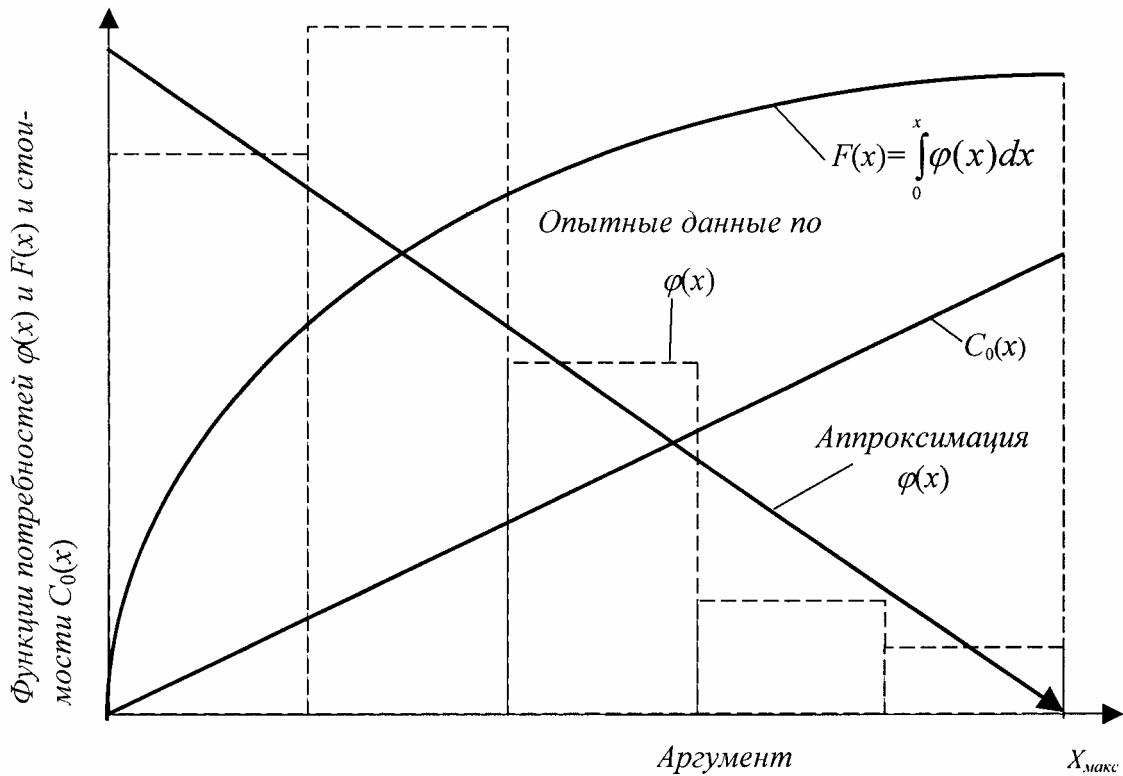


Рис.5.1. Дифференциальная $\phi(x)$ и интегральная $F(x)$ функции потребностей и функция стоимости обслуживания ЭСУД $C_0(x)$

$$\text{Пусть } C_o(x) = ax; \quad (5.14)$$

$$C_p(x) = q_1; \quad (5.15)$$

$$C_s(x) = 0; \quad (5.16)$$

Тогда в зависимости от числа типов оптимальные значения аргументов оказываются такими, как табл.5.1, а суммарные затраты вычисляются по формулам, в правом столбце таблицы.

В таблице: $g_2 = abx^3N$ (5.17)

Пользуясь табл.5.1. при фиксированном числе типов N легко выбрать оптимальные аргументы, а имея данные по g_1 и g_2 и подсчитывая суммарные затраты C при разных N , легко найти то N_i , при котором C минимальна, и выбрать оптимальное N_i .

Таблица 5.1.

Зависимость оптимальных значений аргументов от числа типов суммарных затрат на ЭСУД

Число типов N	Отношение $\frac{x_k}{x_N}$						Суммарные затраты $C = \sum C_p + \sum C_o$
	K						
	1	2	3	4	5	6	
1	1.00						$g_1 + g_2$
2	0.43	1.00					$2g_1 + 0.62g_2$
3	0.27	0.58	1.00				$3g_1 + 0.51g_2$
4	0.19	0.40	0.66	1.00			$4g_1 + 0.46g_2$
5	0.15	0.31	0.49	0.71	1.00		$5g_1 + 0.43g_2$
6	0.12	0.25	0.40	0.56	0.74	1.00	$6g_1 + 0.42g_2$

5.2. Оптимизация решений при выборе типа электронной системы управления двигателями

Основные виды расходов здесь следующие:

1. Расходы, на обслуживание одного автомобиля с ЭСУД в единицу времени – C_S ;
2. расходы, связанные с отсутствием запасных частей (подсистем) для ЭСУД – C_e в обслуживании в определенный промежуток времени;
3. расходы, связанные с пополнением запчастями для ЭСУД, - C_l на обслуживание автомобилей с ЭСУД;
4. расходы, связанные с подбором запасных частей (подсистем), их доработки от других систем; расходы на доработку с уменьшением пропускной способности автосервиса или ремонтного предприятия.

Случай постоянного спроса приводит к следующему решению.

Оптимальное число обслуживаемых мест на автосервисе:

$$n = \sqrt{2\alpha \frac{C}{C_S} \frac{C_S + C_p}{C_p}}, \quad (5.18)$$

где α – спрос, измеряемый числом заявок на обслуживание, выполняемых в единицу времени.

Максимальный объём заявок определяется:

$$S = h \frac{C_p}{C_S + C_p}. \quad (5.19)$$

Суммарные затраты в единицу времени:

$$C = \sqrt{2\alpha C_l C_S \frac{C_p}{C_S + C_p}}. \quad (5.20)$$

Изменение запасных частей происходит следующим образом. Цикл начнётся с того момента, когда запас имеет максимальную величину – S . Затем запас линейно убывает, достигая нуля и затем отрицательной величины ($S - n$) (период дефицита запасных частей), а затем пополняется величиной n и становится равным S . Потом цикл повторяется.

Если дефицит не допускается, то $C_p \rightarrow \infty$ и формулы с (5.18) по (5.20) упрощаются:

$$n = \sqrt{2\alpha \frac{C_l}{C_S}}; \quad (5.21)$$

$$S = n; \quad (5.22)$$

$$C = \sqrt{2\alpha C_l C_S}. \quad (5.23)$$

Поиск решения при случайном спросе дефицитных запчастей (подсистем) ЭСУД рассмотрим на простейшем примере. Пусть α не детерминированная, а случайная величина: известна её плотность распределения вероятностей $f(\alpha)$. Если учитывать только затраты на хранение одной единицы запасных частей ЭСУД без использования информации – C_S , дополнительные затраты при отсутствии заявок на обслуживание – C_p , то решение задачи будет иметь следующий вид. Оптимальная величина максимального запаса (запасных частей ЭСУД) S_0 должна определяться как корень следующего уравнения:

$$\frac{C_p}{C_S + C_p} = \int_0^{S_0} f(\alpha) d\alpha + S_0 \int_{S_0}^{\infty} \frac{f(\alpha)}{\alpha} d\alpha, \quad (5.24)$$

а затраты определяются следующей формулой:

$$C = C_S \int_0^{S_0} (S - \frac{1}{2}\alpha) f(\alpha) d\alpha + C_S \int_{S_0}^{\infty} \frac{S^2}{2\alpha} f(\alpha) d\alpha + C_p \int_{S_0}^{\infty} \frac{(\alpha - S)}{2\alpha} f(\alpha) d\alpha. \quad (5.25)$$

Более или менее удобные формулы удаётся получить, только принимая определённую гипотезу о функции распределения величины α . Так, если считать, что α распределена по закону равной вероятности, то есть:

$$f(\alpha) = \frac{1}{b} \text{ при } 0 \leq \alpha \leq b, \quad (5.26)$$

$$f(\alpha) = 0 \text{ при } \alpha < 0,$$

то формулу (5.24) можно записать следующим образом:

$$\frac{C_p}{C_s + C_p} = \frac{1}{b} \int_0^{S_0} d\alpha + \frac{S_0}{b} \int_{S_0}^b \frac{1}{\alpha} d\alpha,$$

$$\text{откуда: } \frac{C_p}{C_s + C_p} = \frac{S_0}{b} + \frac{S_0}{b} \ln \frac{b}{S_0}. \quad (5.27)$$

Для облегчения решения этого трансцендентного уравнения составим вспомогательную табл. 5.2.

Таблица 5.2.

$\frac{C_p}{C_s + C_p} 10^3$	0	330	520	661	767	847	906	950	979	995	1000
S_0/b	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0

Математическое ожидание расходов может быть получено из совместного решения уравнений (5.25) и (5.26):

$$C = \frac{C_s}{b} \left[\frac{3}{4} S_0^2 + \frac{S_0^2}{2} \ln \frac{b}{S_0} + \frac{C_p}{2bS} \left(\frac{b^2}{2} + \frac{3}{2} S_0^2 - 2bS_0 + S_0^2 \ln \frac{b}{S_0} \right) \right]. \quad (5.28)$$

5.3 Оптимизация процесса технического обслуживания ЭСУД по схеме «затраты - доход» для организации сервиса

Предположим, что имеется E единиц-затрат и W единиц дохода при анализе, электронной системы управления и ее обслуживании (сервисе). Оценке и измерению подлежат потенциальные возможности этой системы: насколько велико число таких реализации? Вероятнее всего система окажется в таком состоянии, в котором она имеет максимальное количество возможных реализации при заданных величинах E и W . И наоборот, исходя из наперёд заданных величин максимального значения, получим множества (линии уровня), со-

стоящих из параметров E и W , соответствующих этим значениям. Один из вариантов оптимального развития ЭСУД можно представить как изменение параметров E и W в таком направлении, когда система наиболее быстро увеличивает максимальные количества своих возможных реализаций. Если просуммировать все максимальные возможности, имеющиеся у системы до текущих значений параметров E и W , то при некоторых условиях сохранения пропорций максимально выраженному росту указанной суммы соответствует степенная зависимость между доходом W и затратами E .

При интерпретации результатов важно учитывать неформальные соображения, которые привели к той или иной информационной характеристике [26].

Пусть система $[(X - \text{затраты}) - (D - \text{оператор}) - (Y - \text{доход})]$ состоит из двух звеньев. (Третье звено D – оператор, то есть **действие** – глагол, условно опустим). **Первое звено** – $(X - \text{затраты})$ – это деньги и работы, на которые они будут израсходованы. **Второе звено** – $(Y - \text{доход})$, – это некоторая сумма денег и продукт (обслуживание), приобретённый на эти деньги (например, запасные части). Работы (или услуги обслуживания) могут быть разными. Это можно учесть, например, с помощью некоторого признака S , заданного для определённости некоторым числом из интервала $[0,1]$.

Обозначим через $B(s)$ и $E(s)$ соответственно количество работ по обслуживанию и количество вложенных в них денежных единиц, приходящихся на единицу признака S . Разобьём отрезок $[0,1]$ точками $S_i (i=1, \dots, n)$ на малые интервалы длины ΔS . Представим, что каждую денежную единицу из $E(S_i)\Delta S$, имеющихся на интервале $(S_i, S_i + \Delta S)$, можно вложить в выполнение одной из $B(s_i)\Delta S$ работ. Число возможных способов будет равно $[B(s_i)\Delta S]^{E(s_i)\Delta S}$. При этом денежные единицы считаем различными, несмотря на их количественное совпадение, что помогает запомнить в какую работу и какая именно единица была вложена.

Пусть далее $U(s)$ и $W(s)$ – соответственно число услуг по обслуживанию ЭСУД и число вложенных в них денежных единиц дохода, приходящиеся на единицу признака S . Представим, что каждую полученную от дохода денежных единиц из $W(S_i)\Delta S$ на интервале $(S_i, S_i + \Delta S)$, можно вложить в одну из $U(S_i)\Delta S$, имеющихся на этом интервале услуг по обслуживанию ЭСУД. Число всевозможных способов будет равно: $[U(S_i)\Delta S]^{W(S_i)\Delta S}$.

Тогда число способов вложения денег в системе из двух звеньев $[(X - \text{затраты}) - (Y - \text{доход})]$ и $\{(Y - \text{доход}) - [E(s) - \text{вложения}]\}$ равно:

$$[B(S_i)\Delta S]^{E(S_i)\Delta S} [U(S_i)\Delta S]^{W(S_i)\Delta S}.$$

Учитывая в положение по всем признакам и их комбинации, найдём число всевозможных способов вложений уже по всем признакам:

$$\prod_{i=1}^n [B(S_i)\Delta S]^{E(S_i)\Delta S} [U(S_i)\Delta S]^{W(S_i)\Delta S}.$$

Полагая, что все эти возможности равновероятны, приходим к величине информации, которую в среднем несёт сообщение о реализации одного из указанных способов вложения в техническое обслуживание ЭСУД:

$$I_i = \ln \left\{ \prod_{i=1}^n [B(S_i) \Delta S] \right\}^{E(S_i) \Delta S} [U(S_i) \Delta S]^{W(S_i) \Delta S}$$

или :

$$I_i = \sum_{i=1}^n [E(S_i) \Delta S] \ln[B(S_i)] + \sum_{i=1}^n [E(S_i) \Delta S] \ln(\Delta S) + \\ + \sum_{i=1}^n [W(S_i) \Delta S] \ln[U(S_i)] + \sum_{i=1}^n [W(S_i) \Delta S] \ln(\Delta S). \quad (5.29)$$

При заданных денежных средств и при малых ΔS только первое и второе слагаемые зависят от вида $B(s)$, $E(s)$, $U(s)$, и $W(s)$. Второе и четвёртое слагаемые одинаковы для всех таких зависимостей с указанным условием на $E(s)$ и $W(s)$.

Далее рассмотрим такую экстремальную задачу:

$$\int_0^1 E(s) ds = E, \quad (5.30)$$

$$\int_0^1 W(s) ds = W, \quad (5.31)$$

$$B(s) \leq B, \quad (5.32)$$

$$U(s) \leq U, \quad (5.33)$$

$$\Phi[E(s), B(s), W(s), U(s)] = \int_0^1 \{E(s) \ln[B(s)] + W(s) \ln[U(s)]\} ds \rightarrow \max, \quad (5.34)$$

где $B(s)$, $E(s)$, $U(s)$, и $W(s)$ – подлежащие определению суммарные почти всюду неотрицательные на отрезке $[0,1]$ функции и такие, что функционал (Φ) имеет конечное значение.

B , E , U , и W – заданные положительные числа. Функционал (Φ) определён по аналогии с конечным пределом при $\Delta S \rightarrow 0$ первого и третьего слагаемых выражения (5.29), который существует, например, при условии непрерывности функции $B(s)$, $E(s)$, $U(s)$, и $W(s)$.

Величина (Φ) имеет смысл некоторой информационной характеристики, описывающей сложность ЭСУД и её потенциальные возможности.

Условие (5.32) задаёт ограничение на величину работ равномерно по всем признакам, а (5.30) ограничивает суммарное количество вложений в обслуживание ЭСУД по всем признакам. Поэтому условие (5.30) можно интерпретировать как возможность концентрировать, например, все денежные вложения в работах, отвечающих некоторому одному признаку S_0 (при этом $E(s) \rightarrow +\infty$ при $s \rightarrow S_0$), а условие (5.32) – как естественное ограничение мощности при производстве работ. Аналогичное замечание можно сделать об условиях (5.31) – (5.33).

Таким образом, решение задачи (5.30) – (5.34) определит такое состояние ЭСУД, в котором она обладает максимальными возможностями при данном уровне денежных средств и мощности реализации работ и услуг по обслуживанию или ремонту.

В силу монотонности логарифма искомое решение задачи (5.30) – (5.34) существует и имеет вид: $B(s) = B$, $U(s) = U$, при почти всех $s \in [0,1]$, а $E(s)$ и $W(s)$ – любые почти всюду неотрицательные функции, удовлетворяющие условиям (5.30) и (5.31). При этом максимальное значение функционала (Φ) равно:

$$I_i = E \ln B + W \ln u. \quad (5.35)$$

Если рассмотреть задачу оптимизации (5.30) – (5.34) с заменой ограничений (5.30) – (5.31) на ограничения типа (5.32) – (5.33), то максимальное значение (5.34) будет равно по-прежнему величине (5.35) в силу монотонности логарифма. Если поменять местами тип ограничений (5.30) – (5.31) и (5.32) – (5.33) [то есть интегральные ограничения на $E(s)$ и $W(s)$ – на равномерное, а равномерные ограничения на $B(s)$ и $U(s)$ – на интегральные], то оказывается, что и в этом случае максимальное значение (5.34) в соответствующей задаче **оптимизации** будет равно величине (5.35) в силу вогнутости логарифма. Заметим также, что можно рассмотреть задачу оптимизации (5.30) – (5.34) с заменой ограничений (5.32) – (5.33) на ограничения типа (5.30) – (5.31). Эта задача не имеет решения: максимизируемый функционал (5.34) можно сделать сколь угодно большим по величине.

Итак, при весьма общих предположениях, максимальное значение информационной характеристики Φ , связанное с максимальными возможностями ЭСУД, зависит только от величин B , E , U , W и не зависит от вида $B(s)$, $E(s)$, $U(s)$ и $W(s)$. При этом максимальная величина информационной характеристики Φ определяется по формуле (5.35).

Таким образом, если предположить, что ЭСУД находится в состоянии с максимальным количеством потенциальных возможностей, определяемых в смысле величины информационной характеристики Φ , то приходим к информационной характеристике (5.35).

5.4. Оптимизация параметров ЭСУД при их изменении

Найдём соотношение между параметрами B , E , U и W при условии, что их изменение приводит к наиболее быстрому возрастанию величины (5.35), то есть максимально быстро увеличивается сложность ЭСУД и её потенциальные возможности. В этом смысле при таком изменении параметров развитие ЭСУД оптимально, и оно определяется направлениями градиентов к поверхно-

стям уровня функции (5.35) как функции переменных B , E , U и W . Отсюда имеем систему дифференциальных уравнений:

$$\overset{\vee}{E} = \ln B, \quad (5.36)$$

$$\overset{\vee}{B} = \frac{E}{B}, \quad (5.37)$$

$$\overset{\vee}{W} = \ln U, \quad (5.38)$$

$$\overset{\vee}{U} = \frac{W}{U}, \quad (5.39)$$

С начальными условиями:

$$E(0) = E_0, \quad B(0) = B_0, \quad W(0) = W_0, \quad U(0) = U_0, \quad (5.40)$$

где B_0 , E_0 , U_0 и W_0 – некоторые заданные числа, большие либо равные единице.

Можно найти решения системы (5.36) – (5.40) в квадратурах и показать, что они обладают свойствами:

$E \rightarrow +\infty$, $B \rightarrow +\infty$, $U \rightarrow +\infty$, $W \rightarrow +\infty$, при $B \sim U$ и $E \sim W$, то есть B – эквивалентна U , а E – эквивалентна W , – предел их ($\lim$) отношения стремиться к единице при соответствующем изменении аргумента.

Таким образом, если параметры системы B , E , U и W меняются так, что информационная характеристика (5.35) максимально быстро изменяет свои значения, то при больших величинах параметров звенья «затраты – вложения» и «доход – вложения» системы начинают функционировать одинаково:

$$B \sim U \text{ и } E \sim W.$$

5.5. Оптимизация параметров ЭСУД с дополнительными условиями при их изменении

Обратимся к случаю или ситуациям, в которых на параметры ЭСУД наложены дополнительные ограничения, то есть существуют некоторые варианты внешнего воздействия на систему или на управление системой.

1. Предположим, что отношение количества работ к количеству услуг (по обслуживанию) есть величина постоянная, равная некоторому положительному значению $k \geq 1/U_0^2$.

Тогда функция (5.35) будет иметь вид:

$$I = \overset{\vee}{E} \ln k U + W \ln U, \quad (5.41)$$

уравнения, определяющие изменения параметров E , U и W системы, при которых значение функции (5.41) изменяется с максимальной скоростью, имеют вид:

$$\overset{\vee}{E} = \ln k U, \quad (5.42)$$

$$\overset{\vee}{U} = \frac{E+W}{U}, \quad (5.43)$$

$$\overset{\vee}{W} = \ln U, \quad (5.44)$$

с начальными условиями:

$$E(0) = E_0, \quad U(0) = U_0, \quad W(0) = W_0. \quad (5.45)$$

Можно найти решения системы (5.42) – (5.45) в квадратурах, что они обладают свойствами $U \rightarrow \infty, E \rightarrow \infty$ и $W \rightarrow \infty$, причём $E \sim W$ (при условии, что количество услуг и работ пропорциональны).

2. Предположим, что отношение количества работ к количеству вложенных в них денег и отношение количества услуг к количеству затраченных на них денег есть постоянные величины, равные соответственно некоторым положительным значениям $k_1 \geq 1/E_0$ и $k_2 \geq 1/W_0$. Тогда функция (5.35) будет иметь вид:

$$I = E \ln k_1 E + W \ln k_2 W. \quad (5.46)$$

Уравнения, определяющие соответствующие изменения параметров, принимают вид:

$$\overset{\vee}{E} = \ln k_1 E + 1, \quad \overset{\vee}{W} = \ln k_2 W + 1 \quad (5.47)$$

с начальными условиями:

$$E(0) = E_0, \quad W(0) = W_0. \quad (5.48)$$

Можно найти решения системы (5.47) – (5.48) в квадратурах и показать, что они обладают свойствами $E \rightarrow \infty, W \rightarrow \infty$, причём $E \sim W$. При этом напомним, что $B = k_1 E, \quad U = k_2 W$.

3. Пусть выполнены предложения п.2. Найдём изменение параметров E и W , при котором с наибольшей скоростью изменяется значение функции:

$$I^o = \int_{E_0}^E x \ln(k_1 x) dx + \int_{W_0}^W y \ln(k_2 y) dy. \quad (5.49)$$

Величина (5.49) – это общая интегральная информационная характеристика, которая каждому состоянию ЭСУД (E, W) ставит в соответствие I^o , являющейся суммой величин потенциальных возможностей всех состояний системы от минимально возможного ($B_{min} = k_1 E = 1, \quad U_{min} = k_2 W = 1;$) до текущего (E, W) . Отметим, что I^o является функцией состояния и в строгом смысле: рассматривая на плоскости параметров (E, W) произвольный путь, соединяющий начальную точку $(1/k_1, 1/k_2)$ и текущую точку (E, W) , получим независимость величины I^o от пути интегрирования в силу потенциальности векторного поля $[E \ln(k_1 E), W \ln(k_2 W)]$.

Уравнения, определяющие соответствующие изменения параметров принимают вид:

$$\overset{\vee}{E} = E \ln k_1 E, \quad \overset{\vee}{W} = W \ln k_2 W, \quad (5.50)$$

с начальными условиями (5.48).

Простое интегрирование системы (5.50) даёт решение в явном виде:

$$E = \frac{1}{k_1} (k_1 E_0)^{\exp(t)}, \quad W = \frac{1}{k_2} (k_2 W_0)^{\exp(t)}. \quad (5.51)$$

Отсюда видно, что переменные E и W связаны некоторой степенной зависимостью

$$W = AE^\gamma, \quad (5.52)$$

$$\text{где } A = \frac{k_1^\gamma}{k_2}, \quad \gamma = k_1 E_0 \ln(k_2 W_0). \quad (5.53)$$

Кроме того, при этом:

$$B = k_1 E, \quad U = A k_2 E^\gamma, \quad \frac{U}{B} = A \frac{k_2}{k_1} E^{\gamma-1}. \quad (5.54)$$

Из выше описанного следует, что оптимальная величина дохода по обслуживанию ЭСУД по затратам – постоянная величина: увеличение затрат на один пункт приводит к изменению дохода на γ пунктов. Такая зависимость появляется в ЭСУД, характеризующихся устойчивым, стабильным функционированием. Это согласуется с условием **оптимального развития при максимально быстром изменении информационной характеристики**, которое быстро можно связать по смыслу с максимально быстрым ростом потенциальных возможностей ЭСУД, ее совершенствованием и надежностью.

Отметим, что после интегрирования (5.49) можно переписать в виде:

$$I^o = \frac{E^2}{2} \ln(k_1 E) + \frac{W^2}{2} \ln(k_2 W) - \frac{E^2}{4} - \frac{W^2}{4} + \frac{1}{4k_1^2} + \frac{1}{4k_2^2}. \quad (5.55)$$

Таким образом, в рамках рассматриваемой модели (X – затраты) – (Y – доход), то есть «вход – выход» в электронной системе управления двигателями, полученные результаты с учетом действий D – операторов, позволяет определить **оптимальные пропорции** между ее составляющими и **оценивать** (потенциальные возможности) состояния ЭСУД в смысле характеристики (5.49), связанной в силу способа ее построения с множеством реализаций системы. От этого зависит тот главный аспект, стоит ли создавать специализированные сервисные предприятия на несколько **хозяйств**, или каждому **хозяйству** строить свои сервисные предприятия (что потребует больших вложений и, возможно, большой срок окупаемости).

6. ОПТИМИЗАЦИЯ ПАРАМЕТРОВ ИЗМЕНЕНИЯ ТЕХНИЧЕСКОГО СОСТОЯНИЯ ЭЛЕКТРОННЫХ СИСТЕМ УПРАВЛЕНИЯ ДВИГАТЕЛЯМИ

Рассмотрим процедуру **оптимизации** параметров технического состояния ЭСУД уже по известным методам классической высшей математики, таких как теория массового обслуживания, теория игр и решений, теории вероятности, теории информации и других методов оптимизации [50].

Например, оптимизацию технического состояния агрегатов и узлов ЭСУД в составе системы - «Автомобиль» обеспечивают не только за счет режима работы подсистем и узлов в заданных эксплуатационных условиях, но и на основе обоснования пробега автомобиля с ЭСУД до предельного состояния и минимальных суммарных затрат на поддержание работоспособности системы в этих условиях. Определить оптимальное [50,52] значение оставшегося пробега автомобиля с ЭСУД до предельного состояния подсистем агрегатов и узлов можно прогнозированием, **статистическими методами прогнозирования** наблюдений и другими экономико-математическими и технико-экономическими методами анализа ЭСУД в составе системы — «Автомобиль» равно как и других технических систем. Качество прогнозирования, уровень его достоверности и точности, уровень затрат на получение прогноза от выбранного метода прогнозирования, принятой математической модели, числа данных и методики определения параметров принятой зависимости. Поэтому необходимо знать основные положения методики определения зависимости. Модель исследования механизма ЭСУД в составе системы — «Автомобиль» представлена на рис. 6.1.

6.1. Определение параметров распределения и линейной оптимизации ЭСУД в составе системы «Автомобиль»

Для практического использования известных или полученных зависимостей необходимо определить численные значения параметров. Наиболее просто определить параметры плотности распределения случайной величины, коэффициенты корреляции [40,42,43,44,45] и корреляционного отношения [42]; более сложно определить параметры линейной параболической зависимости по способу наименьших квадратов; степенные и экспоненциальные зависимости сводятся чаще всего к линейной зависимости математическими преобразованиями.

При известном виде распределения определяют математическое ожидание и дисперсию. Значение параметра, вычисленное на основе ограниченного числа опытов, всегда содержит элемент случайности. Определенные таким образом значения параметров следует называть оценкой.

Среднее значение определяют по зависимости:

$$m = \frac{x_1 + x_2 + \dots + x_i}{n}, \quad (6.1)$$

где x_i - значения отдельных измерений;

n – общее число измерений.

Дисперсия полученных данных:

$$D = \sigma^2 = \frac{\sum (m - x_1)^2 + (m - x_2)^2 + \dots + (m - x_i)^2}{n - 1} \quad (6.2)$$

Среднее квадратическое отклонение единичного измерения $\sigma = \sqrt{D}$, а среднего значения: $\sigma_m = \frac{\sigma}{\sqrt{n}}$ (где n – число измерений).

Определение коэффициента корреляции. Степень соответствия полученных данных прямолинейной зависимости (тесноту связи) оценивают [52,54] с помощью коэффициента корреляции:

$$r = \frac{K_{xy}}{\sigma_x^* \sigma_y^*}, \quad (6.3)$$

где K_{xy} – статистический корреляционный момент;

$\sigma_x^* \sigma_y^*$ – статистические средние квадратические отношения величин x и y .

Оценки коэффициента корреляции с точки зрения точности и значимости определяют по средней погрешности коэффициента корреляции $\varepsilon_r = (1 - r^2)$ при $n > 100$.

Если коэффициент корреляции в 3 раза и более превосходит свою погрешность, то между рассматриваемыми признаками существует корреляционная зависимость.

Определение параметров линейной зависимости по способу наименьших квадратов – это обычно экспериментальные точки на графике имеют случайные отклонения от видимой закономерности. Эти отклонения связаны с неизбежными при всяком опыте погрешностями. Общепринятым при решении подобных задач является способ наименьших квадратов, при котором требование наилучшего согласования кривой и экспериментальных точек сводится к тому, чтобы сумма квадратов отклонений экспериментальных точек от сглаживающей кривой обращалась в минимум.

Если известно, что зависимость линейная, $y = a + bx$, а по результатам эксперимента получена совокупность значений x_i, y_i ($i = 1, 2, \dots, n$), то в соответствии со способом наименьших квадратов параметры a и b должны быть подобраны так, чтобы выполнялось условие:

$$\sum_{i=1}^n [y_i - (bx_i + a)]^2 = \min. \quad (6.4)$$

Чтобы найти значения a и b , при которых левая часть этого выражения будет минимальная, продифференцируем функцию по a и b и приравняем производную к нулю:

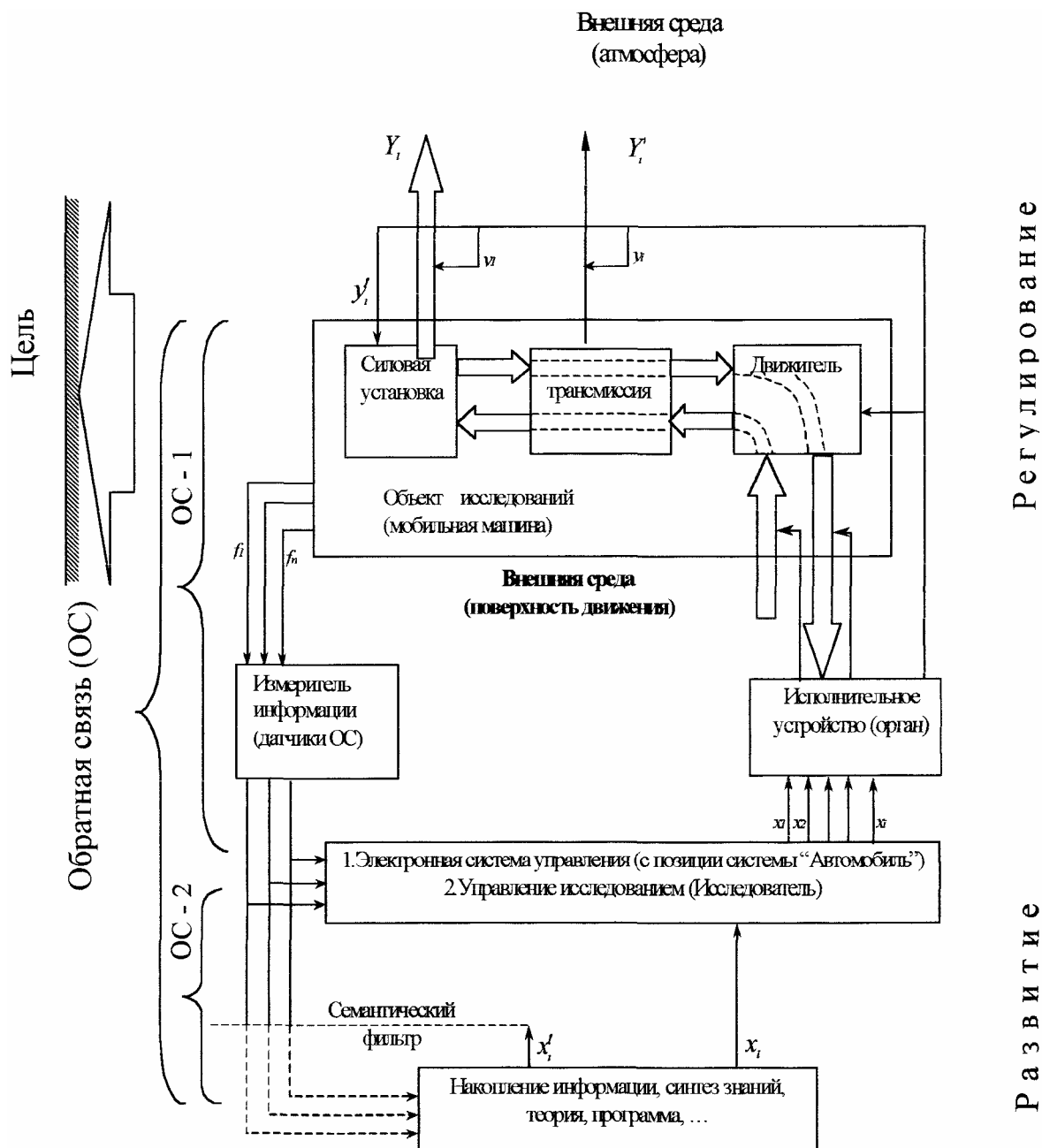


Рис. 6.1. Модель механизма управления исследованием ЭСУД в составе СИСТЕМЫ «Автомобиль»

$$\frac{\partial \varphi}{\partial a} = -2 \sum_{i=1}^n [y_i - (bx_i + a)] = 0; \quad \frac{\partial \varphi}{\partial b} = -2 \sum_{i=1}^n [y_i - (bx_i + a)] x_i = 0. \quad (6.5)$$

где $\frac{\partial \varphi}{\partial a}$, $\frac{\partial \varphi}{\partial b}$ - значения частных производных по параметрам a и b соответственно в точке x_i .

При решении этих уравнений получают более удобную форму записи:

$$b = \frac{K_{xy}^*}{D_x^*}; \quad a = m_y^* - b m_x^*, \quad (6.6)$$

где

$$K_{xy}^* = \frac{\sum_{i=1}^n (x_i - m_x^*)(y_i - m_y^*)}{n} = \frac{1}{n} \sum_{i=1}^n x_i y_i - m_x^* m_y^*;$$

$$D_x^* = \frac{\sum_{i=1}^n (x_i - m_x^*)^2}{n-1} = \frac{n}{n-1} \left(\frac{1}{n} \sum_{i=1}^n x_i^2 - m_x^{*2} \right); \quad (6.7)$$

$$m_x^* = \frac{1}{n} \sum_{i=1}^n x_i; \quad m_y^* = \frac{1}{n} \sum_{i=1}^n y_i.$$

Экспоненциальную зависимость типа $y = y_0 e^{bx}$ сводят к линейной:

$$y' = a' + bx,$$

где $y' = \ln y$; $a' = \ln y_0$.

Методика определения параметров линейной зависимости довольно проста, в частности, координаты одной из точек, через которую пройдет кривая будут m_y и m_x , средние значения y' и пробега x' .

Тогда $\ln y'_0 = \ln m_y - m_x b$. Соответственно, точность прогнозирования изменения технического состояния ЭСУД, а тем самым и оптимизации изме-

нения технического состояния зависит и от методики определения параметров зависимости изменения технического состояния. Наиболее научно обоснована методика определения параметров линейной зависимости по способу наименьших квадратов. Поэтому все остальные формы [54,56] зависимости необходимо сводить к линейной, путем математических методов преобразований.

Экспоненциальную зависимость типа $y = y_0 e^{bx}$ превращают в линейную методом логарифмирования. Сложнее превратить в линейную зависимость типа $y = y_0 - \bar{y}_0 e^{bx}$. Одним из видов этой зависимости является изменение коэффициента технической готовности (выпуска)

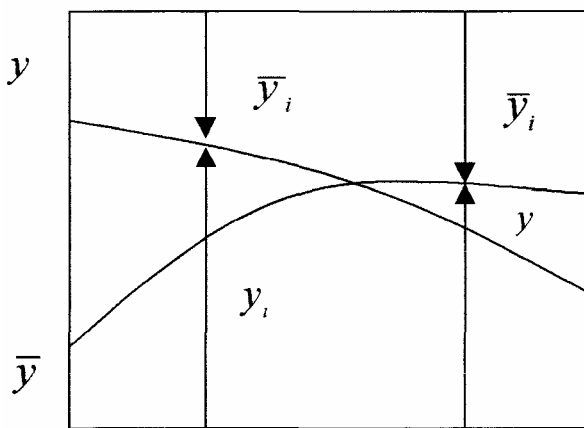


Рис.6.2. Схема определения параметра сложной экспоненциальной зависимости в координатах $y-x$; $\bar{y}-x$ — вспомогательные координаты ($\bar{y}_1, y_1$ — текущие значения y в данных координатах)

$k = 1 - \bar{k}_0 e^{bt}$. По статистическим и экспериментальным данным определяют параметры $\bar{k}_0$ и b , предварительно пересчитав экспериментальные данные при

начале координат с ординатой, равной единице. Для других значений ординат y_0 можно брать произвольно, но ближе к максимальной, больше любого значения результатов статистических наблюдений (рис. 6.2). Например, при расчете параметров закономерности изменения производительности ЭСУД ω за ω_0 принимают производительность при коэффициенте технической готовности, равном единице.

Затем определяют положение результатов изменения в новых координатах $\bar{y}_i - x_i$ с отсчетом $\bar{y}_i$ вниз от вновь принятых координат. В таком случае $\bar{y} + y_i = y_0$. И приведенное сложное уравнение $y = y_0 - \bar{y}_0 e^{bx}$ станет простым общеизвестным $\bar{y} = \bar{y}_0 e^{bx}$, которое легко превратить в линейное. Аналогично поступают с зависимостью $y = y_0(1 - e^{-bx})$, ее сводят к более удобной для определения параметров вида: $\bar{y} = \bar{y}_0 e^{bx}$.

Определение погрешности параметров распределения, линейной зависимости. При использовании зависимости необходимо прежде всего определить параметры, их погрешность. Методика определения погрешности распределения является наиболее простой. Основными начальными параметрами, которые дают первую характеристику распределения, являются математическое ожидание и дисперсия (полученная по известной формуле как средняя арифметическая сумма квадратов разности среднего арифметического и каждой случайной величины), которая характеризует рассеивание результатов наблюдений и эксперимента.

Средняя арифметическая величина является случайной. При увеличении числа наблюдений ее значение сходится по вероятности к математическому ожиданию, которое является величиной постоянной, то есть дисперсия математического ожидания равна нулю. Таким образом, средняя арифметическая величина является случайной и ее колебание характеризуется дисперсией. Дисперсия средней арифметической случайной величины гораздо меньше дисперсии самой случайной величины. В свою очередь дисперсия случайной величины тоже является случайной величиной. Поэтому следует записать не просто m^* и σ^* , а как:

$$m^* \pm \Delta m^*, \sigma^* \pm \Delta \sigma^*, \quad (6.8)$$

где Δm^* и $\Delta \sigma^*$ - погрешность определения m^* и σ^* .

Погрешность случайной величины обычно измеряют в долях от среднего квадратического отклонения. Так, если σ_m^* среднее квадратическое отклонение среднего арифметического, то вместо $\Delta \sigma_i^*$ следует писать $t_\beta \cdot \sigma_{\sigma_i}^*$, в данном случае: $t_\beta = \Delta \sigma_i^* / \sigma_{\sigma_i}^*$. Тогда $m^* \pm \Delta m^* = m^* \pm t_\beta \sigma_m^*$; $\sigma_i^* \pm \Delta \sigma_i^* = \sigma_i^* \pm t_\beta \sigma_{\sigma_i}^*$. Аргументом t_β характеризуют соотношение Δm^* и σ_m^* , $\Delta \sigma_i^*$ и $\sigma_{\sigma_i}^*$ при очень большом числе наблюдений. При ограниченном числе наблюдений t_β следует

использовать t_{β_n} - коэффициент Стьюдента, значение которого определяется по таблице 6.1 в зависимости от числа наблюдений n и значения доверительной вероятности β . Аргумент t_β также определяется по таблице 6.1, но он зависит только от значения доверительной вероятности. Смысл аргумента t_β можно объяснить на примере. Параметрами нормального распределения случайной величины является среднее значение и среднее квадратическое отношение. Практически достаточно точно интервал рассеивания случайной величины может быть принят равным 6σ , то есть все значения случайной величины будут больше $m - 3\sigma$ и меньше $m + 3\sigma$ (теоретически только 0,27% значений случайной величины будут за пределами этого интервала). В интервал $m \pm 2\sigma$ попадает только 0,95 случайных величин, а в интервал $m \pm \sigma$ – лишь 0,68 из них.

Таблица 6.1

Таблица Стьюдента

Число наблюдений, n	Доверительная вероятность, β			Число наблюдений, n	Доверительная вероятность, β		
	0,7	0,9	0,95		0,7	0,9	0,95
2	2,0	3,3	12,7	9	1,1	1,9	2,3
3	1,3	2,9	4,3	10	1,1	1,8	2,3
4	1,3	2,4	3,2	11-14	1,1	1,8	2,2
5	1,2	2,1	2,8	15-16	1,1	1,8	2,1
6	1,2	2,0	2,6	17-27	1,1	1,7	2,1
7	1,1	1,9	2,4	28-40	1,1	1,7	2,0
8	1,1	1,9	2,4	60-120	1,0	1,7	2,0

Таким образом, при доверительной вероятности $0,68t_\beta = 1$. При $\beta = 0,95$ $t_\beta = 2$, а при $\beta = 1,00$ $t_\beta = 3$. Это при бесконечно большом n . При меньшем n и доверительной вероятности $\beta = 0,95$ при числе наблюдений $n = 5$ вместо t_β определяется t_{β_n} , оно, согласно таблице 6.1., равно 2,8; при том же $\beta = 0,95$, но $n = 15$ $t_{\beta_n} = 2,1$. Поэтому и ошибки $\Delta x = 2,86\sigma$ при $n = 5$ и $\Delta x = 2,1\sigma$ при $n = 15$. Так определяется погрешность случайных величин m^* и σ^* параметров нормального распределения. И поэтому: $m^* \pm \Delta m^* = m^* \pm t_{\beta_n} \sigma_m^*$;
 $\sigma^* \pm \Delta \sigma^* = \sigma_i^* \pm t_{\beta_n} \sigma_{\sigma_i}^*$.

Определение погрешности параметров линейной зависимости типа $y = a + bx$ несколько сложнее. В этой зависимости участвуют две случайные величины: среднее m_y^* (заменяющее a) и коэффициент b . Дисперсия среднего в n раз меньше средневзвешенной дисперсии по всем остальным наблюдениям. Сложнее определить погрешность коэффициента b , который имеет сложную зависимость от элементов выборки. Распределение коэффициента b и некото-

рых, связанных с ним величин, наиболее приемлемым из всех известных результатов считается результат, полученный Бартлетом.

Доказано, что:

$$t_{\beta_n} = \frac{b - b_0}{\sigma_y^* \sqrt{1-r}} \sigma_x^* \sqrt{n-2}, \quad (6.9.)$$

где σ_x^*, σ_y^* – среднее квадратическое отклонение x и y ;

r – коэффициент корреляции;

b_0 – истинный коэффициент, имеет распределение Стьюдента с $f = n - 2$ степенями свободы.

Используя t – распределение, можно найти доверительные пределы для коэффициента b , так как остальные величины $\sigma_x^*, \sigma_y^*, r$ и b определяются непосредственно по выборке. При доверительной вероятности $(1-p)$ справедливо неравенство:

$$b - t_{1-p/2} \frac{\sigma_y^* \sqrt{1-r}}{\sigma_x^* \sqrt{n-2}} \leq b_0 \leq b_0 + t_{1-p/2} \frac{\sigma_y^* \sqrt{1-r}}{\sigma_x^* \sqrt{n-2}}. \quad (6.10)$$

После того как найдены доверительные пределы для среднего m_y^* и коэффициента b , можно построить доверительную область, в которой с вероятностью $(1-p)$ лежит истинная линия зависимости. Для построения доверительной области необходимо провести прямые линии через каждую из точек с координатами $m_x^*, m_y^* + \Delta y$ и $m_x^*, m_y^* - \Delta y$ с угловыми

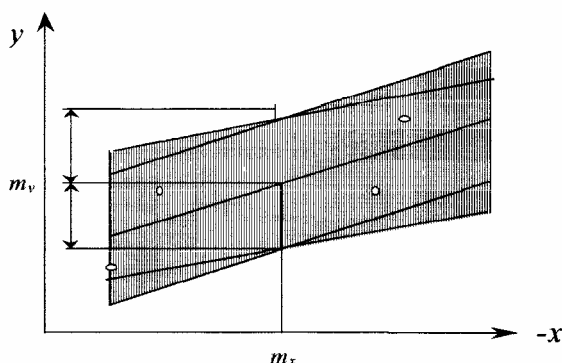


Рис. 6.3. Границы доверительной области линейной зависимости

выми коэффициентами: $b + t_{1-p/2} \frac{\sigma_y^* \sqrt{1-r}}{\sigma_x^* \sqrt{n-2}}$,

$$b - t_{1-p/2} \frac{\sigma_y^* \sqrt{1-r}}{\sigma_x^* \sqrt{n-2}}.$$

Максимальная область, ограниченная этими прямыми, и представляет собой доверительную область (рис. 6.3).

Итак, для определения доверительной области нужно определить Δy и b . Так как

$$\Delta b = t_{1-p/2} \frac{\sigma_y^* \sqrt{1-r}}{\sigma_x^* \sqrt{n-2}}, \text{ то следует определить}$$

коэффициент Стьюдента при заданном числе наблюдений и принятой доверительной вероятности $(1-p)$. Значения коэффициентов $b, \sigma_x^*, \sigma_y^*, r$ определяются по известным формулам.

Таким образом, для всех измерений или результатов нужно указать не только среднее значение, но и доверительный интервал при заданной или принятой доверительной вероятности $m_y + \Delta y$ при принятой β , значение которой обычно

составляет 0,90 или 0,95. Соответственно при определении параметров зависимости $y = a + bx$ указывают $a \pm \Delta a$ и $b \pm \Delta b$ при известной принятой доверительной вероятности, а следовательно, доверительную область, в которой лежит линия истинной зависимости.

Методику определения погрешности рассмотрим на примере. Пусть, например, по результатам износа шатунных шеек двигателей (технологическая система) в зависимости от зазора, в шатунных подшипниках интенсивность изнашивания шеек составила:

Зазор S , мм	10	60	170
Интенсивность изнашивания α , мкм/ 10^4 км.	3,3	5,5	11,3

Необходимо определить параметры α_0 и b линейной зависимости $\alpha = \alpha_0 + bS$.

Решение: Эти параметры можно определить по методу, наименьших квадратов:

$$b = \frac{K_{s_\alpha}^*}{D_s^*}; \alpha_0 = m_\alpha^* - b m_s^*, \quad (6.11)$$

где K_{s_α} - смешанный момент;

D_s^* - дисперсия случайной величины S ;

m_α^* , m_s^* - среднее значение случайных величин α и S ;

α_0 - интенсивность изнашивания, приведенная к началу отчета зазора координат.

По результатам расчета $m_\alpha^* = 6,7 \text{ мкм}/10^4 \text{ км}$; $m_s^* = 80 \text{ мкм}$; $K_{s_\alpha} = 225,3 \text{ мкм}^2 \cdot 10^{-4} \text{ км}$, $D_s^* = 13400/2 = 67 \cdot 10^{-2} \text{ мкм}^2$; $\sigma_s^* = 82 \text{ мкм}$; $D_\alpha^* = 33,9/2 = 16,95 \text{ мкм}^2/10^8 \text{ км}^2$; $\sigma_\alpha^* \cong 4,2 \text{ мкм}/10^4 \text{ км}$; $b = 225/67 \times 100 = 3,36/100 \text{ км}$; $\alpha_0 = 4 \text{ мкм}/10^4 \text{ км}$.

Результаты сведем в таблицу 6.2.

таблица 6.2.

Определение параметров линейной зависимости

S_i	α_i	$S_i - m_s$	$(S_i - m_s)^2$	$\alpha_i - m_\alpha$	$(\alpha_i - m_\alpha)^2$	$(S_i - m_s) \times (\alpha_i - m_\alpha)$	α
10	3,30	-70	4900	-3,4	11,6	238	4,336
60	5,5	-20	400	-1,2	1,44	24	6,02
170	11,3	90	8100	4,6	20,26	414	9,72
-	-	-	-	-	-	-	-
240	20,1	-	13400	-	33,30	676	-

Аналитическое значение α определяется по следующей формуле:
 $\alpha = 4 + 0,0336 \cdot S$. Результаты расчета приводятся в таблице 6.2. вместе с исходными данными. Погрешность параметров:

$$\Delta m_{\alpha} = t_{\beta_n} \sigma_m^*; \Delta b = \pm t_{1-P/2} \frac{\sigma_{\alpha}^* \sqrt{1-r}}{\sigma_s^* \sqrt{n-2}}, \quad (6.12)$$

где σ_{α}^* , σ_s^* - среднее квадратические отклонения величин α и S . В данном случае $r = 0,655$; $n = 3$. При доверительной вероятности $0,9$ $n = 3$, $t = 2,9$, поэтому $\Delta m_{\alpha} = 12,2 \text{ мкм}/10^4 \text{ км}$; $m_{\alpha}^* \pm \Delta m_{\alpha}^* = 6,7 \pm 12,2 \text{ мкм} \cdot 10^4 \text{ км}$; $\Delta b = \pm 4,3 \cdot 0,0176 \text{ 1}/10 \text{ км}$, так как $P = 0,1$; $1-P/2 = 0,95$.

Погрешность при прогнозировании интенсивности изнашивания шатунных шеек может служить характеристикой соответствия полученной зависимости результатам наблюдений. Если за исходные точки принять первые две точки (при 10 и 60 мкм); то $b = (5,5 - 3,30)/(60 - 10) = 0,044$. Интенсивность изнашивания при зазоре 170 мкм составит $\alpha = 5,5 + 0,044(170 - 60) = 5,5 + 0,044 \cdot 110 = 10,34 \text{ мкм}/10^4 \text{ км}$, а по экспериментальным данным $11,3 \text{ мкм}/10^4 \text{ км}$, то есть аналитически полученная интенсивность изнашивания на $1 \text{ мкм}/10^4 \text{ км}$ меньше экспериментальной. Если же сравнить полученные результаты с типовым представлением о постоянстве интенсивности изнашивания, то есть $\alpha = 3,30 \text{ мкм}/10^4 \text{ км}$, то разница составит $11,3 - 3,3 = 8 \text{ мкм}/10^4 \text{ км}$ (в 8 раз больше), то есть погрешность результата в значительной мере зависит от принятой математической модели закономерности. В какой мере погрешность прогнозирования зависит от уровня достоверности принятой закономерности можно судить на примере прогнозирования ресурса до усталостного разрушения деталей.

6.2. Прогнозирование и оптимизация ресурса до усталостного разрушения ЭСУД

В процессе эксплуатации ЭСУД необходимо прогнозировать по первым результатам наблюдений возможный ресурс деталей и узлов системы до усталостного разрушения в заданных эксплуатационных условиях работы. Это позволяет снижать продолжительность исследования, а тем самым и затраты на испытания, увеличить время для принятия решения по ремонтпригодности и ремонту ЭСУД, а также входящих подсистем, деталей и узлов. Прогнозирование ресурса до усталостного разрушения подсистем и узлов ЭСУД проводят по методу линейного суммирования усталостного повреждения.

Комплектующие ЭСУД (детали и подсистемы) испытывают при ступенчатом увеличении нагрузки, начиная с некоторого начального напряжения σ_l , которое повторяется в течении n_l циклов. В процессе испытаний напряжение

увеличивают на $\Delta\sigma$ до σ_2 и на этом уровне испытания продолжают в течении того же числа циклов n_2 то есть $n_2 = n_1$. При каждом увеличении нагрузки число циклов испытания сохраняют постоянным, кроме последнего, значение которого ограничивается моментом разрушения некоторых подсистем ЭСУД. При этом исходят из предположения, что если в эксплуатации подсистема испытала напряжение σ_1 , повторявшееся n_1 раз, то очевидно, что в этих условиях использована только n_1/N_1 часть срока службы N_1 , который подсистема может работать до усталостного разрушения при напряжении σ_1 . Если подсистема в эксплуатации работала с напряжением $\sigma_1, \sigma_2, \sigma_3$ и т.д., повторившимся соответственно n_1, n_2, n_3 и т.д. раз, то, чтобы она не разрушалась ранее установленного срока службы, необходимо соблюдение условия:

$$\frac{n_1}{N_1} + \frac{n_2}{N_2} + \dots \leq 1, \text{ или } \sum_i \left(\frac{n_i}{N_i} \right) \leq 1. \quad (6.13)$$

Однако установлено, что такое условие суммирования является приближенным. В общем виде оно записывается:

$$\sum \left(\frac{n_i}{N_i} \right) = a, \quad (6.14)$$

где a – величина, зависящая от материала детали (комплектующих системы) и условий ее нагружения (загружения).

В приведенной гипотезе [45,53] сделано допущение, которое, по мнению ее сторонников, является очевидным: если при исходном напряжении деталь (система) работает до разрушения N_1 циклов, а она работала только n_1 циклов, использована n_1/N_1 часть ресурса. Применительно к ЭСУД за показатель продолжительности работы удобнее брать пробег; при исходном напряжении σ_1 деталь (система) работает до разрушения l_1 , а фактически отработала только l_{1i} , а соответственно использована l_{1i}/l_1 часть ресурса. В действительности использована значительно меньшая и большая часть ресурса в зависимости от положения участка l_{1i} на кривой в общем пробеге (работе) l_1 с начала эксплуатации.

Изобразим это графически. Методика линейного суммирования усталостных накоплений разрушения была бы справедлива, если интенсивность разрушений в процессе испытаний была постоянной, а площадь усталостных разрушений увеличилась пропорционально продолжительности испытаний и отказ подсистемы наступал при одной и той же площади разрушения (выходе из строя) в абсолютных или относительных (по сравнению с общей рабочей площадью) единицах, как это изображено на рис. 6.4, а.

На рис. 6.4.а линия 1' соответствует интенсивности разрушения при большом исходном разрушении σ_1 , а линия 2' - при малом σ_2 . Только при этих условиях за пробег, например, 0,2 ресурс детали (системы) использован на 20% при работе с исходным напряжением σ_1 . За одинаковые пробеги увеличение площади разрушения (выхода из строя) тоже одинаковое (высота ступеней F_{ii}).

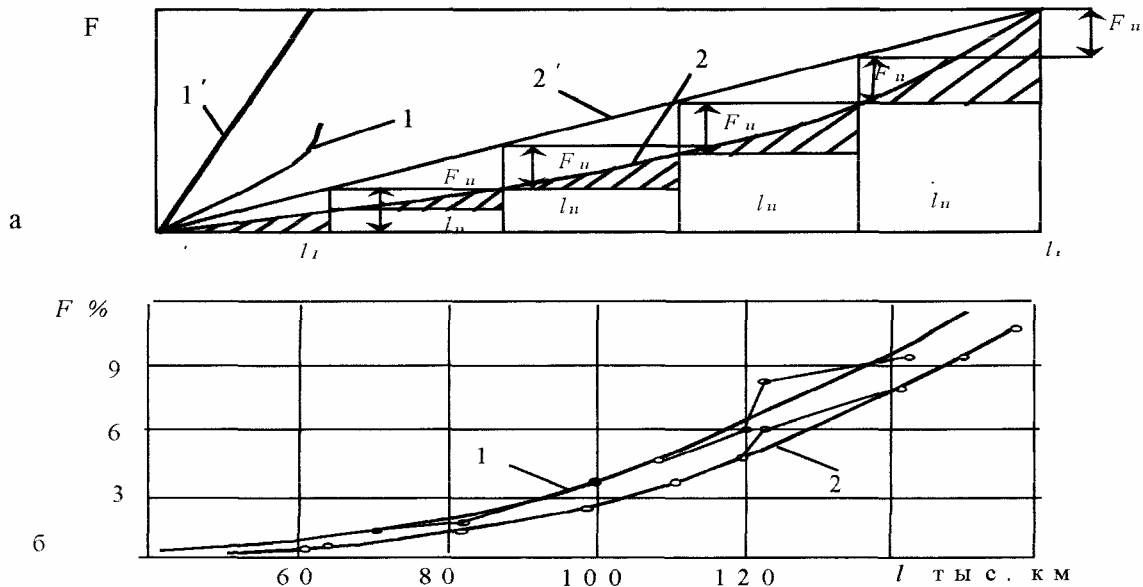


Рис. 6.4. Зависимость площади F усталостных разрушений от продолжительности l испытания:

а - при использовании методики линейного суммирования усталостных накоплений разрушения 1' и 2' соответственно при большом σ_1 и малом σ_2 напряжении в рабочей зоне;

1, 2 - фактические зависимости;

б - для системы впрыска 1 и системы зажигания 2 двигателей с электронной системой от пробега l системы - «Автомобиль»

За этот же отрезок пробега при большом исходном напряжении σ_1 подсистема разрушается. Но если и при исходном напряжении σ_1 продолжительность испытаний взять тоже 0,2 l_i , то в обоих случаях ресурс детали (подсистемы) использован на 0,2! И тогда $\sum \left(\frac{l_{ii}}{l_i} \right) = 1$, где l_i - ресурс детали (подсистемы) при соответствующем исходном напряжении σ_i ; l_{ii} - часть ресурса (продолжительности) испытания.

Фактически интенсивность усталостного разрушения возрастает линейно по мере увеличения площади усталостного разрушения: $\frac{dF}{dl} = a + b\Delta F$
где a - постоянная;

ΔF – увеличение площади разрушения по сравнению с приведенной к началу координат площадью разрушения;

ν – коэффициент интенсификации разрушения.

Поэтому площадь разрушения увеличивается по-разному за равные пробеги: чем дальше от начала координат участок испытаний, тем больше площадь разрушения (заштрихованные участки на рис. 6.4, а) – это **первое**.

Во-вторых, при линейном суммировании усталостных накоплений разрушения по мере эксплуатации предполагается, что разрушение детали (системы) происходит при одинаковой для всех исходных напряжений площади дальнейшего разрушения (долома, выхода из строя).

Фактически площадь остаточного ресурса (долома) пропорциональна исходному напряжению. Поэтому площадь остаточного ресурса (долома) при большом исходном напряжении больше, чем при малом. На рис. 6.4.а площадь остаточного ресурса (долома) при σ_1 при которой происходит долом детали (системы), значительно больше (площадь разрушения меньше), чем при σ_2 .

Это можно записать аналитически. Так, если участок пробега является началом эксплуатации ЭСУД в составе системы – «Автомобиль», то площадь усталостного разрушения в конце пробега l_{ii} запишется:

$$F_n = F_0 e^{bl_n},$$

где F_0 – площадь усталостного разрушения, приведенная к началу координат.

Предельная площадь разрушения

$$F_{\pi} = F_0 e^{bl_{\pi}}, \quad (6.15)$$

где l_n – пробег ЭСУД в составе системы – «Автомобиль» до предельного разрушения детали (системы).

В таком случае:

$$F_n = F_0 e^{bkl_n}, \quad (6.16)$$

где $k = l_n / l_{\pi}$ – доля пробега до предельного разрушения.

Площадь же разрушения составит от F_{π} значительно меньшую часть, если она находится в начале эксплуатации l_{π} .

Поскольку:

$$F_y = F_0 e^{bkl_{\pi}}; F_n = F_0 e^{bl_n},$$

то

$$F_0 = F_{\pi} e^{-bl_{\pi}}; F_n = F_0 e^{bl_n(k-1)}$$

или

$$F_n = F_{\pi} e^{-bl_{\pi}(1-k)}.$$

Из приведенного видно, что связь площади усталостного разрушения с пробегом экспоненциальная и допущение о линейном суммировании усталостных повреждений по мере эксплуатации неправомерно, то есть при $l_n = kl_{II}$ $F_n \neq kF_{II}$. Это подтверждают результаты статистических и эксплуатационных наблюдений за разрушением электронного блока управления ЭСУД и электробензонасоса двигателей системы – «Автомобиль» [48]. Согласно данным статистических наблюдений разрушения электронного блока управления и электробензонасоса двигателей электронной системы управления (рис. 6.4.б), площадь усталостного разрушения увеличивается по мере эксплуатации экспоненциально и поэтому за один и тот же пробег, но в разные периоды эксплуатации, площадь разрушения значительно отличается (при большом пробеге с начала эксплуатации площадь разрушения больше).

Таким образом, исходное положение гипотезы линейного суммирования усталостных повреждений, которую используют при испытаниях на усталость, неправомерно.

Далее, режим испытания на усталость влияет и на предельное состояние поверхности. Так, при испытании [56] образцов установлено, что площади остаточного ресурса (долома) F_D в долях от первоначального сечения образца зависят от первоначального напряжения σ при нагружении (рис. 6.5.)

$$F_D = \frac{\sigma}{\sigma_k},$$

где σ_k – напряжение при разрушении;

σ - исходное напряжение в начале испытания.

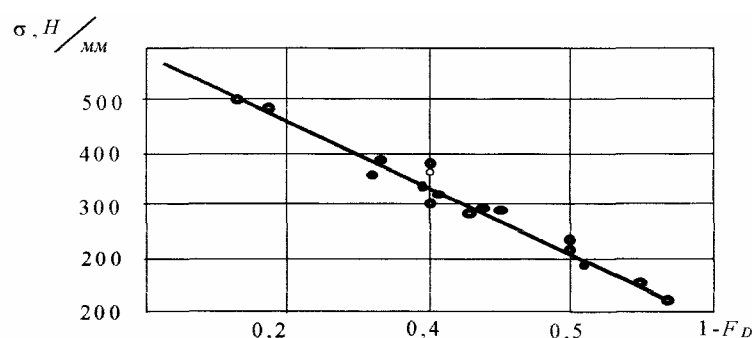


Рис. 6.5. Зависимость предельной степени повреждения $1-F_D$ от максимального напряжения σ цикла для исследованных образцов

Из приведенной зависимости видно, что чем больше исходное напряжение, тем при меньшей площади усталостного разрушения (при большей площади долома) произойдет выход из строя (поломка) детали (системы). Поэтому определить долговечность детали до усталостного разрушения можно только в заданных эксплуатационных условиях.

О влиянии разницы в условиях работы на долговечность подсистем можно судить по кривым износа (рис. 6.6).

Итак, погрешность прогнозирования в установившихся условиях работы зависят от достоверности закономерности изменения технического состояния

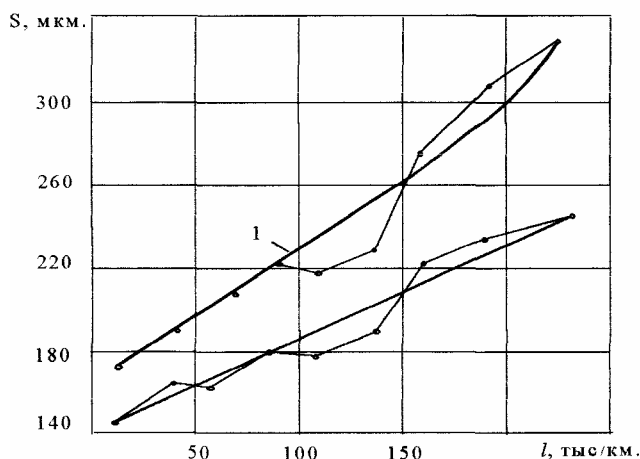


Рис. 6.6. Зависимость подсистемы впрыска S – 1 и системы зажигания – 2 от пробега l ЭСУД в составе автомобиля

сопряжений в процессе эксплуатации ЭСУД (достоверности и числа данных), обоснованности критерия предельного состояния, методики определения параметров закономерности и обоснованности замены подсистем электронных составляющих. И точность, и достоверность прогнозирования должны быть оптимальными в заданных эксплуатационных условиях. Эффективность прогнозирования наработки ЭСУД до технического и профилактического ремонта подсистемы зависит и от эффективности диагностирования технического состояния

ния, от метода определения всей системы и отдельных подсистем. Прогнозирование изменения технико-экономических показателей работы подсистемы в практических условиях возможно при наличии зависимости пробега конкретного критерия состояния, минимального числа данных наблюдения за показателем изменения технического состояния, предельного значения технического состояния, методики определения параметров зависимости. Из перечисленных условий необходимо еще обосновать способ оценки, определения **критерия технического состояния** электронной системы управления двигателями автомобилей (ЭСУД)

6.3. Статистический метод оценки и понятие энтропии технического состояния электронной системы управления двигателями автомобилей

Анализируемый процесс (процедура) оценки технического состояния ЭСУД должен быть описан количественно, иметь количественный измеритель. Вероятностная связь между неисправностями и диагностическими признаками, случайность появления неисправности, информационный характер диагностических признаков позволяют применить математический аппарат теории вероятности, теории информации, основные положения математической статистики и функционального анализа. **Перечисленные** разделы математики и функционального анализа позволяют количественно и качественно описать техни-

ческое состояние ЭСУД в составе системы – «Автомобиль» как объекта диагностирования и анализа. Количественное описание процесса является первейшим условием его **управления и оптимизации**.

Для обоснования численной оценки степени неопределенности можно принять, что к моменту диагностирования и анализа объект (система) может иметь m равновероятных технических состояний, вероятность любого i – го технического состояния $p_i = 1/m$.

Если объект (система) состоит из узлов и агрегатов, каждый из которых может иметь $m_1, m_2, \dots, m_n$ технических состояний, то число технических состояний m объекта (системы) равно произведению чисел технических состояний $m = m_1 m_2 \dots m_n$. Чем меньше число технических состояний, тем меньше степень неопределенности появления каждого технического состояния. При $m = 1$ неопределенность равна нулю, объект (система) может иметь только одно техническое состояние.

С одной стороны численная характеристика [58,60] степени неопределенности объекта (системы) зависит от числа m состояний

$$H(X) = \varphi(m) = \varphi(m_1 m_2 \dots m_n).$$

С другой стороны, вполне можно принять, что степень неопределенности всего объекта равна сумме неопределенности всех его узлов и агрегатов $\varphi(m) = \varphi(m_1) + \varphi(m_2) + \dots + \varphi(m_n)$ или $\varphi(m_1 m_2 \dots m_n) = \varphi(m_1) + \varphi(m_2) + \dots + \varphi(m_n)$.

В математике такими свойствами обладает логарифмическая функция $\log ab = \log a + \log b$. Поэтому:

$H(x) = \log m = \log(m_1 m_2 \dots m_n) = \log m_1 + \log m_2 + \dots + \log m_n$. При равновероятных состояниях отдельно взятое состояние с вероятностью $1/m$ внесет неопределенность в результаты диагностирования и анализа. По значению неопределенности отдельно взятого состояния составит $\frac{1}{m} \log m$, или $-\left(\frac{1}{m}\right) \log\left(\frac{1}{m}\right)$, или $-p_i \log p_i$.

При $p_1 \neq p_2 \neq p_3 \neq \dots \neq p_i$ каждое i – е состояние объекта (системы) вносит неопределенность, равную $-p_i \log p_i$, а общая неопределенность $H(X)$ технических состояний объекта (системы) равна сумме неопределенностей каждого из m состояний:

$$H(x) = -(p_1 \log p_1 + p_2 \log p_2 + \dots + p_i \log p_i), \quad (6.17)$$

где m – число вероятных технических состояний объекта;

p_i – вероятность того, что объект X при диагностировании примет i – е техническое состояние.

Эту сумму $H(X) = -\sum_{i=1}^m p_i \log p_i$ принимают мерой общей неопределенности объекта (системы) диагностирования по возможным техническим состояниям и называют **энтропией** [50,57]. Энтропия является количественной характеристикой уровня неопределенности технического состояния ЭСУД.

Возможность количественно характеризовать уровень неопределенности технического состояния ЭСУД наиболее полно проявляется при анализе статистических данных по надежности подсистем и комплектующих [53].

Узел или агрегат могут находиться в исправном или неисправном состоянии, то есть иметь два технических состояния ($m = 2$). Вероятность исправного состояния $p(l)$ и неисправного $1-p(l)$ зависит от пробега. Таким образом $p_1 = p(l)$, $p_2 = 1-p(l)$.

Поскольку энтропия:

$$H(X) = -\sum_{i=1}^m p_i \log p_i, \text{ то в данном случае:}$$

$$H(X) = -\sum_{i=1}^m p_i \log p_i = -(p_1 \log p_1 + p_2 \log p_2) = \quad (6.18)$$

$$= \{[p(l) \log p(l)] + [1-p(l)] \log [1-p(l)]\}$$

При средней наработке вероятности отказа и безотказности работы узла или агрегата одинаковы и поэтому энтропия наибольшая, (табл. 6.3):

Энтропию измеряют в двоичных единицах (битах). За **бит** принимают энтропию подсистемы (системы), который с одинаковой вероятностью может иметь одно из двух возможных состояний и потому:

$$p_i = \frac{1}{m} = \frac{1}{2}; \quad H(X) = -\left[\left(\frac{1}{2}\right) \log_2 \frac{1}{2} + \left(\frac{1}{2}\right) \log_2 \frac{1}{2}\right] = -\log_2 \frac{1}{2} = \log_2 2 = 1.$$

Таблица 6.3.

Вероятность безотказной работы системы (подсистемы)

Показатель	Пробег, /, тыс. км.						
	15	25	35	45	55	65	75
Число отказов подсистемы (системы) в интервале	1	8	17	45	16	11	2
Накопленное число системы объекта в конце i интервала	1	9	26	71	87	98	100
Вероятность безотказной работы	0,99	0,91	0,74	0,29	0,13	0,02	0,00
Энтропия	0,2104	0,4364	0,8268	0,8687	0,5574	0,3988	0

Число возможных состояний подсистем ЭСУД и вероятность каждого из них устанавливают по результатам статистических данных наблюдений за изменением технического состояния подсистем в процессе эксплуатации и расчетам по таблице 6.4. при помощи таблицы 6.5.

Таблица 6.4.

Расчет энтропии

p_i	$\bar{p}_i = 1 - p_i$	$-p_i \log p_i$	$-\bar{p}_i \log \bar{p}_i$	$-(p_i \log p_i + \bar{p}_i \log \bar{p}_i)$
0,99	0,01	0,144	0,0664	0,2104
0,91	0,09	0,1238	0,3126	0,4364
0,74	0,26	0,3215	0,5053	0,8268
0,29	0,71	0,5179	0,3508	0,8687
0,13	0,87	0,3826	0,1748	0,5574
0,02	0,98	0,1128	0,286	0,3988

Кривые изменения энтропии и информации, полученные на основе статистических данных работы ЭСУД до отказа, повторяют кривую плотности распределения пробега до отказа. При равной вероятности каждого из состояний значение энтропии максимальное, необходимость диагностирования и анализа особенно велика.

Таким образом, значение энтропии характеризует уровень неопределенности технического состояния системы (подсистемы) в количественной форме, целесообразность использования углубленного диагностирования даже простого узла.

Энтропия сложной системы. Электронная система управлением двигателями - это сложная система, состоит из подсистем, узлов, подборок, и деталей, равно как и любой объект подчиняется принципу иерархичности построения по входимости и уровням иерархии, а техническое состояние этой системы взаимосвязано с техническим состоянием подсистем ЭСУД, а также агрегатов и узлов всей системы «Автомобиль», от чего зависит и значение энтропии этой системы, агрегатов при известной энтропии каждого узла и детали (подсистемы). Взаимосвязанными являются диагностический показатель и структурный параметр. При диагностировании по диагностическому показателю необходимо определить техническое состояние узла, сопряжения, значение их структурного параметра. Поэтому по изменению энтропии диагностического признака можно определить изменения энтропии структурного параметра технического состояния.

Энтропия обладает свойством аддитивности, то есть когда несколько независимых систем объединяются в одну, то энтропии складываются объединением двух систем X и Y с возможными состояниями $x_1, x_2, \dots, x_n; y_1, y_2, \dots, y_n$ называют сложную систему $\{X, Y\}$ состояние которой $\{x_i, y_j\}$ представляет со-

бой все возможные комбинации состояний x_i, y_j систем X и Y . Число возможных состояний объединения $\{X, Y\}$ равно $m \times n$, то вероятность, что $\{X, Y\}$ будет в состоянии $\{x_i, y_j\}$ обозначают через P_{ij} : $P\{x \sim x_i, y \sim y_j\} = P_{ij}$. В общем виде энтропия сложной системы равна сумме произведений вероятностей всех возможных ее состояний на их логарифм с обратным знаком:

$$H(X, Y) = - \sum_{i=1}^m \sum_{j=1}^n P_{ij} \log P_{ij} \quad (6.19)$$

Энтропию сложной системы вычисляют в соответствии с теоремой умножения вероятностей для независимых событий $P(X, Y) = P(X)P(Y)$.

При определении энтропии систем, составленной из зависимых элементов, простое сложение неприменимо. В этом случае определяют условную вероятность $P\left(\frac{y_j}{x_i}\right)$ систем Y , когда X приняла конкретное значение x_i . Таким объединением взаимосвязанных систем является диагностирование, когда при конкретном значении диагностического признака x_i система X ; структурный параметр (система Y) составит y_j .

Условная энтропия системы Y при условии, что система X находится в состоянии x_i , $H\left(\frac{Y}{x_i}\right) = - \sum_{j=1}^n P\left(\frac{y_j}{x_i}\right) \cdot \log P\left(\frac{y_j}{x_i}\right)$, или:

$$H\left(\frac{Y}{x_i}\right) = \sum_{j=1}^n \eta \left[P\left(\frac{y_j}{x_i}\right) \right] \quad (6.20)$$

Условная энтропия системы Y зависит от того, какое состояние x_i приняла система X . Если вероятность этого состояния p_i , то среднюю или полную условную энтропию определяют по общепринятому методу:

$$H\left(\frac{Y}{X}\right) = - \sum_{i=1}^m p_i H\left[P\left(\frac{Y}{x_i}\right)\right]. \quad (6.21)$$

Если $H\left(\frac{Y}{x_i}\right)$ записать в развернутом виде:

$$H\left(\frac{Y}{x_i}\right) = - \sum_{j=1}^n p_i \sum_{j=1}^n P\left(\frac{y_j}{x_i}\right) \cdot \log P\left(\frac{y_j}{x_i}\right) \text{ или после внесения } p_i \text{ под знак}$$

$$\text{второй суммы } H\left(\frac{Y}{X}\right) = - \sum_{i=1}^m \sum_{j=1}^n p_i P\left(\frac{y_j}{x_i}\right) \cdot \log P\left(\frac{y_j}{x_i}\right), \text{ или } H\left(\frac{Y}{X}\right) = \sum_{i=1}^m \sum_{j=1}^n p_i \eta \left[P\left(\frac{y_j}{x_i}\right) \right].$$

Таблица 6.5

Данные расчета

x	$\log x$	x	$\log x$	x	$\log x$
1.	0,00000	35.	5,21928		
2.	1,00000	36.	5,16993	69.	6,10852
3.	1,58496	37.	5,20945	70.	6,12928
4.	2,00000	38.	5,24793	71.	6,14975
5.	2,32193	39.	5,28540	72.	6,16992
6.	2,058496	40.	5,32193	73.	6,18982
7.	2,80735	41.	5,35755	74.	6,20945
8.	3,00000	42.	5,39232	75.	6,22882
9.	3,16993	43.	5,42626	76.	6,24793
10.	3,32193	44.	5,45943	77.	6,26679
11.	3,45943	45.	5,49185	78.	6,28540
12.	3,58496	46.	5,52356	79.	6,30378
13.	3,70044	47.	5,55459	80.	6,32193
14.	3,80735	48.	5,58496	81.	6,33985
15.	3,90689	49.	5,61471	82.	6,35755
16.	4,0000	50.	5,64386	83.	6,37504
17.	4,08746	51.	5,67242	84.	6,39232
18.	4,16993	52.	5,70044	85.	6,40933
19.	4,24793	53.	5,72492	86.	6,42626
20.	4,32193	54.	5,75489	87.	6,44294
21.	4,39232	55.	5,78136	88.	6,45943
22.	4,45943	56.	5,80735	89.	6,47573
23.	4,52356	57.	5,83289	90.	6,49185
24.	4,58496	58.	5,85798	91.	6,50779
25.	4,64836	59.	5,88264	92.	6,52356
26.	4,70044	60.	5,90689	93.	6,53916
27.	4,75489	61.	5,93074	94.	6,55459
28.	4,80735	62.	5,95420	95.	6,56986
29.	4,85798	63.	5,97728	96.	6,58496
30.	4,90689	64.	6,00000	97.	6,59991
31.	4,95420	65.	6,02237	98.	6,61471
32.	5,00000	66.	6,04439	99.	6,62936
33.	5,04439 1	67.	6,06609	100.	6,64386
34.	5,08746	68.	6,08746		

По теореме умножения вероятностей $p_i P\left(\frac{y_j}{x_i}\right) = P_{y_j}$, и поэтому

$$H\left(\frac{Y}{X}\right) = -\sum_{i=1}^m \sum_{j=1}^n P_{y_j} \cdot \log P\left(\frac{y_j}{x_i}\right). \text{Значение } H\left(\frac{Y}{X}\right) \text{ называют полной условной}$$

энтропией системы Y относительно X . Она характеризует степень неопределенности системы Y оставшуюся после того, как состояние системы X полностью определилось [56].

6.4. Оптимизация затрат и критерии качества текущего ремонта и обслуживания ЭСУД

Основные положения теории массового обслуживания используют для оптимизации затрат и планировании текущего ремонта и обслуживания ЭСУД с учетом особенностей этой системы.

Для характеристики степени совершенства организации технического обслуживания и ремонта ЭСУД в составе системы - «Автомобиль», необходимо знать количественную зависимость расходов на поддержание работоспособности этой системы. Эффективность любой организации технического обслуживания и ремонта системы можно проверить как экспериментально, так и при помощи теории массового обслуживания.

Предметом теории массового обслуживания является количественная сторона организации процесса массового обслуживания [60], или количественная оценка качества функционирования системы массового обслуживания. Поступившую в систему обслуживания заявку (например, на ремонт ЭСУД в составе системы — «Автомобиль») называют требованием на обслуживание, а удовлетворение этого требования (заявки) - **обслуживанием**. Под обслуживающим аппаратом, каналом и устройством понимают то, что способно обслуживать в данный момент только одно требование. Так как ремонт или техническое обслуживание ЭСУД выполняют не на одном посту, а на нескольких, то такую группу однородных постов (обслуживающих аппаратов, каналов) называют **обслуживающей системой**. **Качество** функционирования обслуживающей системы при заданной организации ее работы и количестве поступающих требований можно оценить при помощи теории массового обслуживания. Задачей теории массового обслуживания является отыскание функциональных зависимостей величин, которые характеризуют качество функционирования действия обслуживающей системы от характеристики входящего потока, от параметров одного канала обслуживания и от способов организации всей обслуживающей системы. Показателями **эффективности** обслуживания принимают вероятность простоя системы (каналов), среднее число ЭСУД в составе системы -«Автомобиль» в ожидании текущего ремонта или профилактики. (Процесс функционирования системы массового обслуживания при текущем ремонте

или профилактике ЭСУД представляет собой случайный процесс). Случайные процессы в системах массового обслуживания автомобилей с ЭСУД представляет собой процессы с непрерывным пробегом (временем), поэтому для теоретического расчета пропускной способности системы следует знать характеристику случайного процесса, закономерность распределения пробега (времени) обслуживания и характер потока заявок. **В потоке событий**, события являются однородными, так как взяты, например, автомобили, которые поступают только в текущий ремонт или на профилактику только по одному. Для системы по текущему ремонту или профилактике ЭСУД в составе системы – «Автомобиль» самым характерным является простейший поток событий: стационарный, без последствия, ординарный. В теории массового обслуживания считают, что большинство потоков являются простейшими, а если и отличаются от простейших, то с вполне удовлетворительной точностью результатов поток любой структуры можно заменить простейшими с такой же плотностью.

Для простейшего потока число автомобилей с ЭСУД, поступающих в текущий ремонт или на профилактику на участке пробега Δl , распределенного по закону Пуассона с математическим ожиданием $m = \lambda \Delta l$, где λ - среднее число автомобилей, поступающих в текущий ремонт за единицу пробега (плотность потока).

Признаками пуассоновского распределения вероятности поступления k автомобилей с ЭСУД при плотности потока отказов λ является равенство по значению математического ожидания и дисперсии $m = D$; $\sigma^2 = m$. Поэтому число требований на обслуживание:

$$N_e = m \pm t_{\beta} \sigma, \text{ или } N_e = m \pm t_{\beta} \sqrt{m},$$

где t_{β} – коэффициент Стьюдента.

Максимальная вероятность поступления автомобилей с ЭСУД в ремонт при $k = m = \lambda \Delta l$. Вероятность того, что за пробег Δl поступит k автомобилей с ЭСУД:

$$p_k(\Delta l) = \frac{(\lambda \Delta l)^k}{k!} \cdot e^{-\lambda \Delta l}, \quad (6.22)$$

а вероятность того, что не поступит ни одной системы ($k = 0$):

$$p_0(\Delta l) = e^{-\lambda \Delta l}. \quad (6.23)$$

Итак, известна Δl , следует знать и закон, функцию распределения длины пробега l_i между поступления систем – «Автомобиль» с ЭСУД в текущий ремонт или на профилактику:

$$F(\Delta l) = p(l_i < \Delta l),$$

где l_i - случайная величина, пробег между двумя соседними поступлениями систем – «Автомобиль» с ЭСУД в простейшем потоке.

Сравнительно, легче определить вероятность, противоположного события: $1 - F(\Delta l) = P(l_i > \Delta l)$, то есть вероятность того, что на участке длиной Δl , который начинается в момент поступления автомобиля с ЭСУД в текущий ремонт или на профилактику, больше в текущий ремонт или на профилактику не поступит. Вероятность такого события $p(\Delta l) = e^{-\lambda \Delta l}$, откуда

$$1 - F(\Delta l) = e^{-\lambda \Delta l}; F(\Delta l) = 1 - e^{-\lambda \Delta l} (\Delta l > 0) \quad (6.24)$$

Плотность приведенного распределения

$$f(\Delta l) = \lambda e^{-\lambda \Delta l}, \quad (6.25)$$

Закон распределения с такой плотностью называется показательным, а λ – его параметром. Пробег обслуживания Δl имеет показательное распределения, то

$$f(\Delta l) = \mu e^{-\mu \Delta l}, \quad (6.25')$$

где μ – среднее число автомобилей с ЭСУД; а $f(\Delta l)$ – плотность распределения пробега обслуживания.

При показательном законе распределения дисперсия пробега и математическое ожидание равны по величине: $D = m_l; \sigma^2 = m_l$.

Следовательно, общий пробег обслуживания или ремонта с учетом дисперсии

$$l_i = m_l \pm t_\beta \sqrt{m_l},$$

где m_l – средний пробег обслуживания.

При таком законе распределения пробега обслуживания будет справедливо утверждение: если в какой-то момент пробега происходит обслуживание заявки, то закон распределения оставшегося пробега обслуживания не зависит от того, какой пробег обслуживания уже продолжался и так далее. При математическом описании процесса массового обслуживания ЭСУД в составе системы – «Автомобиль» ремонтом или обслуживанием исходят из показательного закона распределения пробега до ремонта или обслуживания и потерянного пробега за время текущего ремонта или обслуживания ЭСУД; за показатели эффективности обслуживания принимают вероятность простоя системы и среднее число ЭСУД в составе системы «Автомобиль» в очереди на ремонт

или обслуживание, а качество ремонта или обслуживания зависит от убытков из-за простоя системы «Автомобиль» с ЭСУД в ремонте или на обслуживании и его ожидания.

Показатели эффективности обслуживания (вероятность простоя системы, среднее число автомобилей в очереди) устанавливают при помощи зависимостей теории массового обслуживания [53]. Систему массового обслуживания ЭСУД в составе системы «Автомобиль» текущим ремонтом или обслуживанием (профилактикой) называют **системой с ожиданием**, так как ЭСУД, которой необходимо сделать текущий ремонт или техническую профилактику, не может работать пока этот ремонт или техническое обслуживание не произведены. Как отмечалось выше, состояние одноканальной системы с ожиданием бесконечное (хотя и счетное) множество. Возможные состояния:

x_0 – канал свободен (не занят);
 x_1 – канал занят;
 x_2 – канал занят, одна заявка стоит в очереди;
;
 x_{1+i} – канал занят, i заявок стоит в очереди.

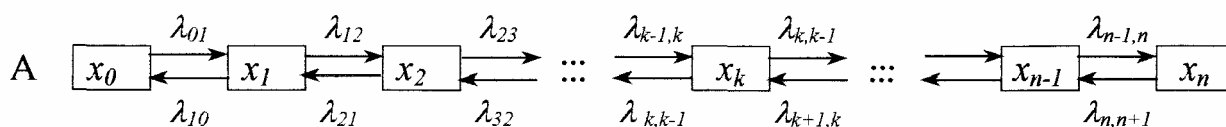
Возможные состояния канала и их вероятности:

$$p_0(l), p_1(l), p_2(l), \dots, p_{1+i}(l), \quad (6.26)$$

Очевидно, для любого пробега

$$\sum_{k=0}^{1+i} p_k(l) = 1 \quad (6.27)$$

Аналитическое определение $p_0(l), p_1(l), p_2(l), \dots, p_{1+i}(l)$ состояние системы массового обслуживания зоны текущего ремонта или обслуживания удобно начинать с более понятного и удобного для математического описания (рис. 6.7 а) графика, используемого для так называемой системы «гибели-размножения». Особенность этого графика в том, что все состояния системы можно вытянуть в одну цепочку, в которой каждое из средних состояний ($x_1, x_2, \dots, x_{n-1}$) связано прямой и обратной связями (стрелками) с каждым из соседних соединений (правым и левым), а крайние состояния x_0, x_n , – только с одним соседним состоянием.



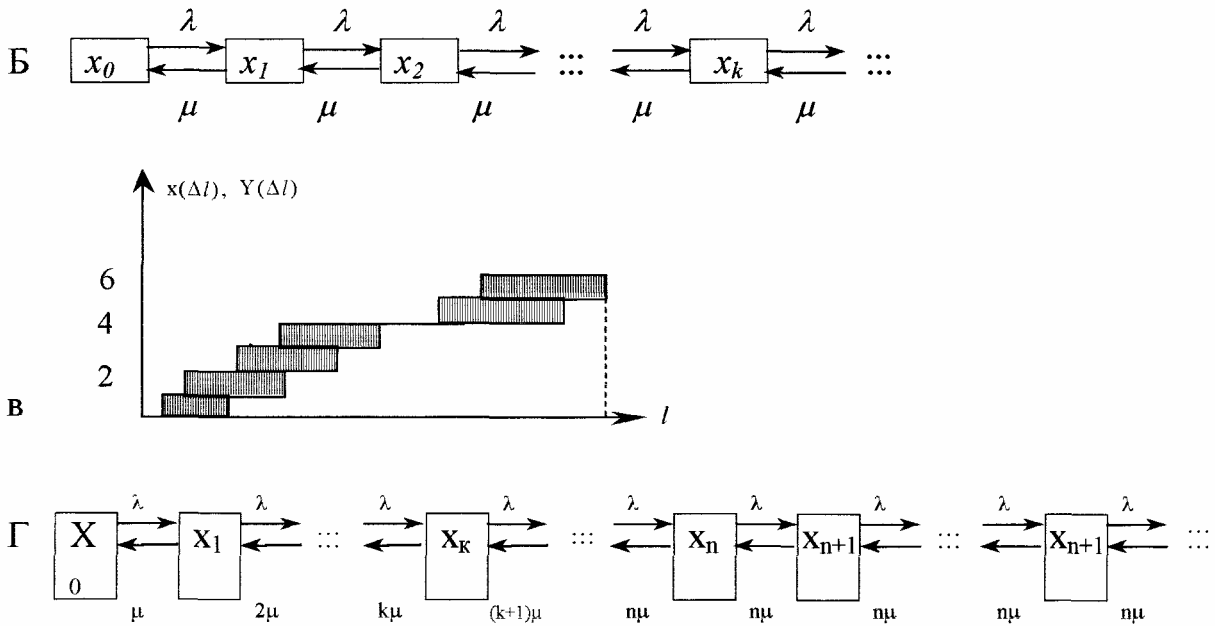


Рис. 6.7. Состояние системы массового обслуживания:

а – для системы «гибели-размножения» при конечном числе состояний;

б – одноканальный с неограниченной очередью;

в – при изменении в процессе работы системы массового обслуживания числа заявок прибывших $X(\Delta l)$ и покинувших $Y(\Delta l)$ систему массового обслуживания до момента Δl ;

г – n – канальной с неограниченной очередью.

Как отмечалось выше, все потоки событий, приводящие систему по стрелкам графа, простейшие. Прямо по графу состояний можно написать систему линейных алгебраических уравнений, где с одной стороны **финальная** вероятность данного состояния p_k , умноженная на суммарную интенсивность всех потоков ведущих из данного состояния, а с другой стороны – сумма произведений интенсивностей всех потоков, входящих в k -е состояние, на вероятность тех состояний, из которых эти потоки исходят.

[**Примечание:** финальными вероятностями состояний называют пределы вероятностей, если они существуют и не зависят от начального состояния, системы при неограниченной продолжительности, при $l \rightarrow \infty$]. Состояние системы массового обслуживания рассмотрим на следующем **примере**:

Пользуясь графиком рис. 6.7 а, составим и решим алгебраические уравнения для финальных вероятностей состояний (их существования вытекает из того, что из каждого состояния можно перейти в каждое другое, и число состояний конечно). Для первого состояния x_0

$$\lambda_{01} p_0 = \lambda_{10} p_1 \quad (6.28)$$

Для состояния x_1 :

$$(\lambda_{12} + \lambda_{10}) p_1 = \lambda_{01} p_0 + \lambda_{21} p_2$$

С учетом формулы (6.28) последнее равенство упрощается: $\lambda_{21} p_1 = \lambda_{21} p_2$; аналогично записывается $\lambda_{23} p_2 = \lambda_{32} p_3$;

и в общем виде: $\lambda_{k-1,k} p_{k-1} = \lambda_{k,k-1} p_k$,

где $0 \leq k \leq n$.

Финальные вероятности $p_0, p_1, \dots, p_n$ можно определить при решении системы уравнений:

$$\left. \begin{aligned} \lambda_{01} p_0 &= \lambda_{10} p_1; \\ \lambda_{12} p_1 &= \lambda_{21} p_2; \\ &\dots\dots\dots; \\ \lambda_{k-1,k} p_{k-1} &= \lambda_{k,k-1} p_k; \\ &\dots\dots\dots; \\ \lambda_{n-1,n} p_{n-1} &= \lambda_{n,n-1} p_n. \end{aligned} \right\} \quad (6.29)$$

С учетом нормированного условия:

$$p_0 + p_1 + p_2 + \dots + p_n = 1 \quad (6.30)$$

Из первого уравнения системы (6.29):

$$p_1 = \left(\frac{\lambda_{01}}{\lambda_{10}} \right) p_0. \quad (6.31)$$

Из второго уравнения системы (6.29) с учетом формулы (6.31) получим:

$$p_2 = \frac{\lambda_{12}}{\lambda_{21}} p_1 = \frac{\lambda_{12} \lambda_{01}}{\lambda_{21} \lambda_{10}} p_0. \quad (6.32)$$

Аналогично:

$$p_3 = \frac{\lambda_{23} \lambda_{12} \lambda_{01}}{\lambda_{23} \lambda_{21} \lambda_{10}} p_0. \quad (6.33)$$

и в общем виде:

$$p_k = \frac{\lambda_{k-1,k} \lambda_{12} \lambda_{01}}{\lambda_{k,k-1} \lambda_{21} \lambda_{10}} p_0, \quad 1 \leq k \leq n \quad (6.34)$$

Как видно из формулы (6.34), в числителе – произведение всех интенсивностей, стоящих у стрелок (см. рис. 6.7), ведущих слева направо (с начала и до данного состояния x_k), интенсивностей поступления заявок на обслуживание, а в знаменателе – произведение всех интенсивностей, стоящих у стрелок, ведущих справа налево (с начала и до данного состояния x_k) – интенсивностей обслуживания заявок.

В соответствии с нормированным условием (6.30):

$$p_0 \left(1 + \frac{\lambda_{01}}{\lambda_{10}} + \frac{\lambda_{12}\lambda_{01}}{\lambda_{21}\lambda_{10}} + \dots + \frac{\lambda_{n-1,n}\lambda_{12}\lambda_{01}}{\lambda_{n,n-1}\lambda_{21}\lambda_{10}} \right) = 1.$$

Следовательно:

$$p_0 = \left(1 + \frac{\lambda_{01}}{\lambda_{10}} + \frac{\lambda_{12}\lambda_{01}}{\lambda_{21}\lambda_{10}} + \dots + \frac{\lambda_{n-1,n}\lambda_{12}\lambda_{01}}{\lambda_{n,n-1}\lambda_{21}\lambda_{10}} \right)^{-1}. \quad (6.35)$$

Все остальные вероятности выражены через p_0 [см. формулы (6.31) – (6.34)]. Коэффициенты при p_0 в каждой из формул представляет собой последовательные члены ряда, стоящего после единицы в формуле (6.35), значит, вычисляя p_0 , уже найдены все эти коэффициенты.

В рассматриваемой системе массового обслуживания зоны текущего ремонта или техобслуживания поступают заявки на обслуживание с интенсивностью λ , а интенсивность потока обслуживаний μ . Теоретически число состояний ничем не ограничено, граф состояний имеет вид, показанный на рис. 6.7,б.

Эта схема «гибели-размножения» с бесконечным числом состояний. По всем стрелкам поток заявок с интенсивностью λ переводит систему слева направо, а справа налево – поток обслуживающий с интенсивностью μ . Финальные вероятности для таких систем массового обслуживания существует только в том случае, если система не перегружена, то есть если соотношение числа поступающих заявок и возможности системы массового обслуживания, при $\mu > \lambda$, $\rho = \lambda/\mu < 1$.

Формулы для финальных вероятностей в схеме «гибели-размножения» выведены только для случая конечного числа состояний, но ими можно воспользоваться и для бесконечного числа состояний. В том случае число слагаемых формулы (6.35) буде бесконечным:

$$p_0 = \left[1 + \frac{\lambda}{\mu} + \left(\frac{\lambda}{\mu} \right)^2 + \dots + \left(\frac{\lambda}{\mu} \right)^k + \dots \right]^{-1} = (1 + \rho + \rho^2 + \dots + \rho^k + \dots)^{-1}. \quad (6.36)$$

Ряд в формуле (6.36) представляет собой геометрическую прогрессию. Известно, что при $\rho < 1$ ряд сходится – это бесконечно убывающая прогрессия со знаменателем ρ .

В общем виде сумма прогрессии $\frac{a - z_q}{1 - q}$ (где a – первый член прогрессии; z – последний член прогрессии; q – знаменатель).

Для рассматриваемого нами случая $a = 1$, $q = \rho$, $z = \rho^k$.

При $k \rightarrow \infty$, $\rho < 1$, $\rho^k \rightarrow 0$

$$\sum_{k=1}^{\infty} \rho^k = \frac{1 - \rho^{\infty}}{1 - \rho} = \frac{1}{(1 - \rho)}.$$

Поэтому $1 + \rho + \rho^2 + \dots + \rho^k + \dots = \frac{1}{1 - \rho},$

$$P_0 = 1 - \rho \quad (6.37)$$

[Примечание: при $\rho = 1$ система массового обслуживания может справиться только с регулярным потоком заявок и если время обслуживания равно интервалу между заявками, чего практически обеспечить невозможно. Эти параметры всегда случайны и поэтому очередь на обслуживание будет увеличиваться при $\rho = 1$.

При большом колебании времени ремонта или техобслуживания около математического ожидания необходимо обеспечить высокий уровень работ на постах и достаточный резерв рабочих и оборудования, практически невозможно наладить устойчивый, хорошо сбалансированный процесс работы на потоке].

Вероятности $p_1, p_2, \dots, p_k$ находят по формулам:

$$p_1 = \rho p_0, \quad p_2 = \rho^2 p_0, \dots, \quad p_k = \rho^k p_0, \dots,$$

оттуда с учетом (6.37):

$$p_1 = \rho(1 - \rho), \quad p_2 = \rho^2(1 - \rho), \dots, \quad p_k = \rho^k(1 - \rho). \quad (6.38)$$

Как видно, вероятности $p_0, p_1, \dots, p_k$ образуют геометрическую прогрессию со знаменателем ρ , математическая вероятность p_0 (вероятность, что канал будет вообще свободен). Как бы ни была нагружена система с очередью при $\rho < 1$, самое вероятное число заявок в системе будет 0.

Иногда следует знать не только p_0 и m (целое число заявок в очереди), но и среднее число заявок в системе, среднюю продолжительность прерывания заявки в системе l_c , в том числе в очереди l_{0r} при предельном стационарном режиме. В этих условиях среднее число заявок на обслуживание и убывающих за единицу продолжительности, интенсивность λ одинаковы.

Обозначим: $X(\Delta t)$ - число заявок прибывающих в систему массового обслуживания до момента Δt ; $Y(\Delta t)$ число заявок поступивших в систему массового обслуживания до момента Δt . Обе функции случайные, изменяются на единицу: по одному автомобилю поступает в зону текущего ремонта и по од-

ному убывает. Вид функций $X(\Delta l)$ и $Y(\Delta l)$ показан на рис. 6.7,в; обе линии ступенчатые; верхняя $X(\Delta l)$, нижняя $Y(\Delta l)$.

Разница между ними в любой момент – это число заявок, находящихся в системе массового обслуживания:

$$Z(\Delta l) = X(\Delta l) - Y(\Delta l).$$

Средне число заявок, находящихся в системе массового обслуживания за очень большой промежуток l , равно интегралу от функции $Z(\Delta l)$ на этом промежутке, деленному на длину интервала l :

$$M_c = \frac{1}{l} \int_0^l Z(\Delta l) dl \quad (6.39)$$

Графически этот интеграл состоит из совокупности прямоугольников высотой в единицу, с длиной, равной продолжительности пребывания заявки в системе (заштрихованные площади на рис. 6.7,в).

Обозначим эти продолжительности пребывания заявки $\Delta l_1, \Delta l_2, \dots$. Если под конец промежутка некоторые прямоугольники войдут в заштрихованные площади не полностью, то при достаточно большом интервале этим можно пренебречь. Таким образом можно считать:

$$\int_0^l Z(\Delta l) dl = \sum_i \Delta l_i \quad (6.40)$$

Тогда в соответствии с формулой (6.39)

$$M_c = \frac{1}{l} \int_0^l Z(\Delta l) dl = \frac{1}{l} \sum_i \Delta l_i. \quad (6.41)$$

Разделив и умножив правую часть формулы (6.41) на интенсивность λ , получим:

$$M_c = \frac{1}{l\lambda} = \sum_i \Delta l_i \lambda.$$

Но $l\lambda$ – это среднее число заявок, пришедших в период l . Если разделить сумму всех Δl_i на среднее число заявок, то получим среднюю продолжительность пребывания заявок в системе l_c .

Таким образом, $M_c = \lambda l_c$ отсюда:

$$l_c = \frac{1}{\lambda} M_c \quad (6.42)$$

Формула (6.42) справедлива для всех видов системы массового обслуживания потока, заявок, распределения продолжительности обслуживания и дисциплины обслуживания. Ее суть: средняя продолжительность пребывания заявки в системе равна среднему числу заявок в системе, деленному на интенсивность потока заявок. Аналогично получается зависимость средней продолжительности пребывания заявки в очереди $l_{оч}$ и от среднего числа заявок в очереди m :

$$l_{оч} = \frac{1}{\lambda} m. \quad (6.43)$$

При выводе достаточно вместо нижней линии, на рис. 6.7, взять функцию $U|\Delta|$ - число заявок, ушедших до момента не из системы, а из очереди (если заявка пришедшая в систему, не становится в очередь, сразу идет на обслуживание, можно все же считать, что она становится в очередь, но находится в ней нулевое время).

Среднее число заявок в системе массового обслуживания можно определить, если известны вероятности $p_0, p_1, \dots, p_k$ наличия в системе соответственно 0, 1, 2, ..., k заявок:

$$M_c = 0p_0 + 1p_1 + 2p_2 + \dots + kp_k + \dots = \sum_{k=1}^{\infty} kp_k \quad (6.44)$$

(сумма берется не от 0 до ∞ , а от 1 до ∞ так как нулевой член равен нулю). Или с учетом формулы (6.38)

$$M_c = \sum_{k=1}^{\infty} k\rho^k (1-\rho)$$

Если же вывести за знак суммы $\rho(1-\rho)$: $M_c = \rho(1-\rho) \sum_{k=1}^{\infty} k\rho^{k-1}$, то под знаком суммы остается $k\rho^{k-1}$ которая является производной по ρ от выражения ρ^k . следовательно,

$$M_c = \rho(1-\rho) \sum_{k=1}^{\infty} \frac{d}{d\rho} \rho^k,$$

или

$$M_c = \rho(1-\rho) \frac{d}{d\rho} \sum_{k=1}^{\infty} \frac{d}{d\rho} \rho^k. \quad (6.45)$$

Сумма же в формуле (6.45) – это сумма бесконечно убывающей геометрической прогрессии с первым членом и знаменателем ρ , она равна $\frac{\rho}{1-\rho}$, а ее производная $\frac{1}{(1-\rho)^2}$. С учетом всего формула (6.45) значительно упрощается

$$M_c = \frac{\rho}{(1-\rho)}. \quad (6.46)$$

Средняя продолжительность пребывания заявки в системе в соответствии формулой (6.42):

$$l_c = \frac{M_c}{\lambda} = \frac{\rho}{\lambda(1-\rho)} \quad (6.47)$$

Из находящихся в системе массового обслуживания заявок часть уже обслуживается, а остальные ожидают своей очереди. Среднее число заявок в системе в очереди по правилу сложения математических ожиданий равно среднему числу заявок в системе минус среднее число заявок под обслуживанием. Под обслуживанием может быть или одна заявка, или ни одной. Вероятность того, что канал занят:

$$p_{зан} = 1 - p_0 = \rho. \quad (6.48)$$

Поэтому среднее число заявок под обслуживанием:

$$M_{об} = 1 \cdot \rho = \rho. \quad (6.49)$$

Следовательно, среднее число заявок в очереди $m = M_c - \rho = \frac{\rho}{1-\rho} - \rho$. И окончательно:

$$m = \frac{\rho^2}{(1-\rho)} \quad (6.50)$$

В соответствии с формулой (6.43) средняя продолжительность пребывания заявки в очереди:

$$l_{оч} = \frac{\rho^2}{\lambda(1-\rho)}. \quad (6.51)$$

Таким образом определяются характеристики эффективности работы одноканальной системы массового обслуживания применительно в зоне текущего ремонта или технического обслуживания при бригадном обслуживании ЭСУД в составе системы – «Автомобиль». **Многоканальная система массового обслуживания с неограниченной очередью.** Зона текущего ремонта или технического обслуживания обычно имеет несколько постов. Характеристики

эффективности многоканальной системы массового обслуживания с неограниченной очередью определяют, как для одноканальной, но несколько сложнее. Состояния нумеруют также по числу заявок в системе:

- x_0 – в системе массового обслуживания заявок нет; все каналы свободны;
- x_1 – занят один канал, остальные свободны;
- x_2 – занято два канала, остальные свободны;
-
- x_k – занято k каналов, остальные свободны;
-
- x_n – занято n каналов, очереди нет;
- x_{n+1} – занято n каналов, одна заявка стоит в очереди;
-
- x_{n+i} – занято n каналов, i заявок стоят в очереди.

Граф состояний многоканальной системы массового обслуживания с неограниченной длиной очереди показан на рис. 6.7, г, он является схемой «гибели-размножения»: у стрелок указаны интенсивности потоков событий. Из x_0 в x_1 систему приводит приход заявки, поток заявок с интенсивностью λ . Этот же поток переводит систему из любого левого в соседнее правое состояние, как показано верхними стрелками. Переход системы из правого состояния в левое происходит несколько сложнее: из состояния x_1 (работает один канал) в x_0 система сможет перейти, если канал x_1 производит μ обслуживаний в единицу времени; это и указано у стрелки $x_1 \rightarrow x_0$. Из состояния x_2 (работают два канала) система может перейти в x_1 только в том случае, если закончил обслуживание либо первый, либо второй канал. Суммарная интенсивность их обслуживания равна 2μ , что и отмечено у соответствующей стрелки. Суммарный поток обслуживающий тремя каналами имеет интенсивность 3μ , k каналами – $k\mu$, это отмечено у соответствующих нижних стрелок.

Существование финальных вероятностей возможно при $\rho/n < 1$. В этом случае по формулам (6.34) и (6.35) для схемы «гибели-размножения» можно найти эти финальные вероятности. В выражении для p_0 будет стоять ряд членов, содержащих факториалы плюс сумма бесконечно убывающей геометрической прогрессии со знаменателем ρ/n . Суммируя ее, найдем:

$$\left. \begin{aligned} p_0 &= \left[1 + \frac{\rho}{1!} + \frac{\rho^2}{2!} + \dots + \frac{\rho^n}{n!} + \frac{\rho^{n+1}}{n!(n-\rho)} \right]^{-1} \\ p_1 &= \frac{\rho}{1!} p_0, \dots, p_k = \frac{\rho^k}{k!} p_0, \dots, p_n = \frac{\rho^n}{n!} p_0; \\ p_{n+1} &= \frac{\rho^{n+1}}{n \times n!} p_0, \dots, p_{n+i} = \frac{\rho^{n+i}}{n^i n!} p_0 \end{aligned} \right\} \quad (6.52)$$

Из характеристик эффективности системы массового обслуживания легче всего определить среднее число занятых каналов $M_k = \lambda/\mu = \rho$ (эта формула справедлива для любой системы массового обслуживания с неограниченной очередью).

Среднее число заявок в очереди определить проще, чем в системе, это можно сделать по формуле: $m = \sum_{i=1}^{\infty} i p_{n+1}$.

После соответствующих преобразований с дифференцированием ряда (как и в одноканальной системе массового обслуживания) среднее число заявок в очереди:

$$m = \frac{\rho^{n+1} p_0}{n \times n! \left(1 - \frac{\rho}{n}\right)^2} \quad (6.53)$$

Суммируя среднее число заявок в очереди и под обслуживанием (оно же среднее число занятых каналов $M_k = \rho$), можно определить и среднее число заявок в системе:

$$M_c = m + \rho \quad (6.54)$$

Если же полученные значения для средних значений числа заявок в очереди и системе разделить на интенсивность поступлений заявок, можно получить среднюю продолжительность пребывания заявки соответственно в очереди и в системе:

$$l_{oc} = \frac{1}{\lambda} m; \quad l_c = \frac{1}{\lambda} M_c. \quad (6.55)$$

Таким образом, **показатели эффективности** обслуживания (то есть вероятность простоя p_0 зоны текущего ремонта или технического обслуживания среднего числа m ЭСУД в составе систем – «Автомобиль» в ожидании ремонта или технического обслуживания) зависят от приведенной плотности.

6.5. Оптимизация пропускной способности зоны текущего ремонта и обслуживания ЭСУД в составе системы «Автомобиль»

Критерием эффективности расчета [50,52] принимают показатели или параметры, которые количественно оценивают соответствие полученных результатов заранее запланированным. В данном случае такими критериями могут быть продолжительность простоя в ожидании ремонта или технического обслуживания или длина очереди, число свободных постов, экономический эффект.

Для реализации поставленной цели каждым процессом, в том числе изменением технического состояния и использованием ЭСУД, необходимо управлять. Управление называется **оптимальными**, если **изменение состояния**

объекта (переход его из одного состояния в другое) происходит по наиболее выгодному, оптимальному пути.

Оптимизацию параметров пропускной способности осуществляют по экономическому критерию: или по минимуму затрат, или по максимуму удельного расхода, то есть дохода от транспортной работы высвобождаемых из очереди ЭСУД в составе систем - «Автомобиль» на единицу расходов при работе системы.

При проведении оптимизации должна минимизироваться только одна из величин, остальные могут быть, если это необходимо, ограничены каким-либо условиями. Результаты расчета оптимальной пропускной способности (числа постов) по минимуму удельных затрат и по максимуму удельного дохода совпадают очень редко.

При текущем ремонте или техническом обслуживании ЭСУД в составе системы - «Автомобиль» в затраты на поддержание системы в технически исправном состоянии входят заработная плата рабочим: слесарям, электрикам, электроникам и другим исполнителям, амортизация зданий, сооружений, стоимость запасных частей (обоснованность их замены), убытки от простоя системы - «Автомобиль» с ЭСУД в ожидании и в процессе текущего ремонта или технического обслуживания, постов в ожидании поступления системы - «Автомобиль» с ЭСУД на текущий ремонт или техническое обслуживание, накладные расходы и др.

Минимум затрат на текущий ремонт или техническое обслуживание ЭСУД в составе системы - «Автомобиль» в заданных эксплуатационных условиях и пробеге сначала эксплуатации конкретной модели системы - «Автомобиль» с ЭСУД можно обеспечить в результате оптимизации числа рабочих в зоне текущего ремонта или технического обслуживания, оборотных запасных подсистем, узлов и деталей (агрегатов) на складе, совмещения замен деталей, из-за которых уже произошел отказ, и других деталей с большей вероятностью отказа ко времени замены отказавшей, оптимизацией пробега до капитального ремонта или проведения планового технического осмотра и обслуживания.

При оптимальном планировании обеспечивают, с одной стороны, минимальные суммарные убытки от простоя системы - «Автомобиль» с ЭСУД в ожидании ремонта или технического обслуживания из-за отсутствия или поста, или необходимых деталей, агрегатов (подсистем), узлов, а с другой - убытки из-за простоя поста системы массового обслуживания в ожидании текущего ремонта ЭСУД в составе системы - «Автомобиль» и от хранения невостребованных до определенного времени запасных частей или подсистем ЭСУД.

В отличие от метода планирования текущего ремонта или технического обслуживания конкретной ЭСУД в составе системы - «Автомобиль» планируют и потребное число рабочих, постов обслуживания, оборотных агрегатов (подсистем) и технических узлов, деталей для выполнения текущего ремонта или технического обслуживания с минимальными затратами. Поскольку при нерегулируемом поступлении системы - «Автомобиль» с ЭСУД в текущий ре-

монта или на техническое обслуживание, ее простой или поста – события случайные, поэтому расчет оптимального числа обслуживающего персонала, оборотных агрегатов (подсистем), запасных частей ведется на основе теории массового обслуживания.

При методе обслуживания, например, бригадным подрядом всю зону технического обслуживания и текущего ремонта можно рассматривать как один пост, а весь обслуживающий персонал – универсалами, то есть взаимозаменяемыми.

В соответствии с общепринятой методикой при оптимальном параметре суммарные удельные затраты минимальны. Применительно к рассматриваемому случаю необходимо принять такое соотношение между числом поступивших на обслуживание заявок и пропускной способностью канала, чтобы средние удельные убытки от простоя канала и системы - «Автомобиль» с ЭСУД в очереди в сумме были минимальны по сравнению с другими вариантами соотношения.

Для этого необходимо иметь прежде всего данные об убытке за единицу времени при простое системы (или канала) a_1 и ЭСУД в составе системы - «Автомобиль» a_2 .

При известных вероятности простоя поста $\rho_0 = 1 - \rho$ и среднем числе простаивающих систем - «Автомобиль» с ЭСУД $m = \frac{\rho^2}{1 - \rho}$ необходимо определить

$$\rho, \text{ при котором } \rho_0 a_1 + m a_2 = (1 - \rho) a_1 + \frac{\rho^2}{1 - \rho} a_2 = \min.$$

Значение ρ можно определить из уравнения

$$\frac{d}{d\rho} \left[(1 - \rho) a_1 + \frac{\rho^2}{1 - \rho} a_2 \right] = 0.$$

При дифференцировании зависимости необходимо учитывать, что производная отношения $\left(\frac{u}{v} \right)' = \frac{v du - u dv}{v^2}$. Поэтому при дифференцировании последнего уравнения:

$$-a_1 + a_2 \left[\frac{(1 - \rho) 2\rho + \rho^2}{(1 - \rho)^2} \right] = 0,$$

$$\text{а затем } -a_1 + 2(a_2 + a_1)\rho - \rho^2(a_1 + a_2) = 0.$$

Таким образом получено квадратичное уравнение типа $ax^2 + bx + c = 0$, в котором значение неизвестного:

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

Следовательно, оптимальное значение приведенной плотности

$$\rho = \frac{-2(a_1 + a_2) \pm \sqrt{4(a_1 + a_2)^2 - 4(a_1 + a_2)a_1}}{-2(a_1 + a_2)}$$

или $\rho = 1 \pm \sqrt{\frac{a_2}{a_1 + a_2}}$. Так как $\rho < 1$, то $\rho = 1 - \sqrt{\frac{a_2}{a_1 + a_2}}$.

Оптимальная пропускная способность:

$$\mu = \frac{\lambda}{\left(1 - \sqrt{\frac{a^2}{a_1 + a_2}}\right)}.$$

Оптимальная приведенная плотность зависит от соотношения убытка из-за простоя канала обслуживания и системы – «Автомобиль» с ЭСУД. Если убытки от простоя системы – «Автомобиль» с ЭСУД, например, составляют 31 руб./ч, то тарифная ставка рабочего 4,56 руб./ч оптимальная приведенная плотность $\rho = 1 - \sqrt{\frac{31}{4,56 + 31}} = 0,053$, то есть оптимальная пропускная способность в $1/0,053 = 18,8$ раза больше числа заявок на ремонт или техническое обслуживание. Однако убытки от простоя канала обслуживания нельзя сводить только к простоям рабочих. **Во-первых**, число рабочих всегда больше единицы; **во-вторых** необходимо учитывать убытки от простоя оборудования. В связи с этим увеличивается соотношение a_1/a_2 убытков от простоя поста a_1 и автомобилей a_2 оптимальная приведенная плотность увеличивается.

Оптимальная приведенная плотность с увеличением убытка из-за простоя канала по сравнению с убытком от простоя автомобиля с ЭСУД в очереди плавно нарастает (рис. 6.8.):

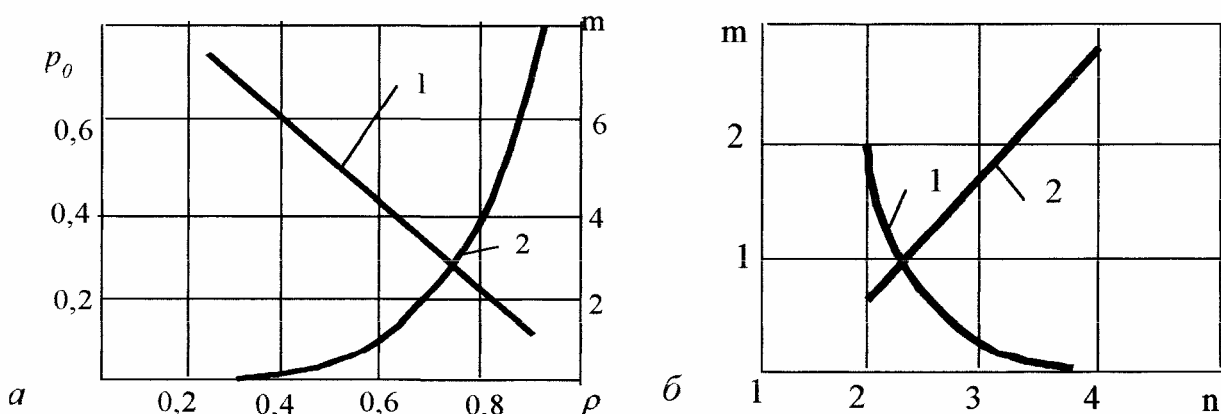


Рис. 6.8. Зависимость показателей эффективности обслуживания p_0 и m :

а – от удельной плотности поступления заявок на канал обслуживания: 1 – вероятность p_0 простоя канала; 2 – среднее число m автомобилей с ЭСУД в ожидании обслуживания;

б – от числа каналов n обслуживания при $\rho = 3/2$; 1 – среднего числа m автомобилей в ожидании обслуживания; 2 – каналов в ожидании автомобилей с ЭСУД на обслуживание.

При $a_1 \gg a_2$ значение ρ приближается к единице, но остается все-таки меньше единицы. Такой характер зависимости ρ от a_1/a_2 вполне объясним: по мере увеличения ρ , p_0 убытки от простоя канала убывают пропорционально, а убытки от простоя автомобиля, нарастают по степенной зависимости $\rho^2(1-\rho)$.

Таким образом, для расчета оптимальной пропускной способности зоны текущего ремонта в практических условиях, кроме среднего значения числа требований ЭСУД в составе системы – «Автомобиль» на ремонт, необходимо знать убытки от простоя канала обслуживания (зоны текущего ремонта или технического обслуживания) системы – «Автомобиль» с ЭСУД в ожидании ремонта (или техобслуживания) или их соотношения. При известных параметрах увеличения трудоемкости текущего ремонта или технического обслуживания на единицу пробега в процессе эксплуатации можно принять меры по корректированию пропускной способности зоны в зависимости от пробега с начала эксплуатации.

Из вышеизложенного можно сделать следующий очередной вывод: поскольку простой поста (зоны обслуживания) зависит от приведенной плотности, то при оптимальном значении приведенной плотности ρ и вероятность простоя зоны обслуживания будет оптимальной:

$$p_0 = 1 - \rho = \sqrt{\frac{a_2}{a_1 + a_2}}. \text{ Следовательно, простой рабочих увеличивается, если}$$

убытки от простоя системы – «Автомобиль» с ЭСУД растут быстрее убытков от простоя зоны обслуживания текущего ремонта или обслуживания.

Оптимальная очередь ЭСУД в составе системы – «Автомобиль» на ремонт или техническое обслуживание составляет:

$$m_{opt} = \frac{\rho_{opt}^2}{1 - \rho_{opt}} = \frac{1}{p_{0opt}} \rho^2 = \frac{1}{p_{0opt}} \left[\left(1 - \sqrt{\frac{a^2}{a_1 + a_2}} \right) \right]^2 = \frac{1 - 2\sqrt{\frac{a^2}{a_1 + a_2}} + \frac{a_2}{a_1 + a_2}}{\sqrt{\frac{a_2}{a_1 + a_2}}} = \frac{1}{p_{0opt}} - 2 + p_{0opt},$$

$$\text{или же } m_{opt} = \frac{1 - 2p_{0opt} + p_{0opt}^2}{p_{0opt}} = \frac{(1 - p_{0opt})}{p_{0opt}}.$$

Если же число каналов n обслуживания два и более, пропускная способность каждого канала μ , число ЭСУД в составе систем – «Автомобиль», поступающих за единицу времени (пробега) на ремонт λ , $\rho = \lambda/\mu$, то:

$$p_0 = \frac{1}{\sum_{k=0}^n \frac{\rho^k}{k!} + \frac{\rho^{n+1}}{n!(n-\rho)}};$$

$$p_k = \frac{\frac{\rho^k}{k!}}{\sum_{k=0}^n \frac{\rho^k}{k!} + \frac{\rho^{n+1}}{n!(n-\rho)}}, \quad 0 \leq k \leq n$$

$$p_{k+i} = \frac{\frac{\rho^{n+i}}{n!n^i}}{\sum_{k=0}^n \frac{\rho^k}{k!} + \frac{\rho^{n+1}}{n!(n-\rho)}}, \quad i \geq 0; \quad k = n+i,$$

где k - ЭСУД в составе систем - «Автомобиль» на обслуживании в системе обслуживания; i - число ЭСУД в ожидании ремонта или технического обслуживания;

$$m = \frac{n \times n! \left(1 - \frac{\rho}{n}\right)^2}{\sum_{k=0}^n \frac{\rho^k}{k!} + \frac{\rho^{n+1}}{n!(n-\rho)}}.$$

В этом случае $\rho = \lambda/\mu$ при $\left(\lambda/n\mu\right) < 1$.

Минимальное число постов текущего ремонта берут на 0,2 и более приведенной плотности. Более устойчивая работа зоны при соотношении приведенной плотности и числа постов $0,65 \div 0,85$. Приведенной плотностью называют среднее число требований поступивших за средний пробег обслуживания одного требования, ремонта одной ЭСУД в составе системы - «Автомобиль».

Оптимальное число постов n_{opt} определяется способом перебора, то есть определяется: $m_n = \varphi(n)$, $m_i = \varphi(n)$ и в конечном счете определяется $a_1 m_n + a_2 m_i = \varphi(n)$, находя n_{opt} , то есть n , при котором:

$$a_1 m_n + a_2 m_i = \min,$$

где a_1 - убытки от простоя поста в единицу пробега;

m_n - среднее число простаивающих незанятых каналов;

a_2 - убытки от простоя системы - «Автомобиль» с ЭСУД в ожидании ремонта или технического обслуживания.

Среднее число незанятых каналов зависит от вероятности p_k поступления в ремонт k ЭСУД в составе систем «Автомобиль». Если каждый канал, пост обслуживает, например, две ЭСУД, то из общего числа n постов, на которые с вероятностью p_k поступит k таких систем (ЭСУД), будут простаивать

$\left(n - \frac{k}{2}\right)p_k$ постов. Если пропускная возможность поста μ , то число простаивающих постов:

$$m = \sum_{k=0}^n \left(n - \frac{k}{\mu}\right)p_k.$$

Посты будут простаивать, если $n > k/\mu$, поэтому при определении среднего числа простаивающих постов m_n необходимо суммировать все положительные значения $\left(n - \frac{k}{\mu}\right)p_k$.

Все вышеизложенное рассмотрим на числовом **примере**:

Определить оптимальное число постов зоны текущего ремонта станции технического обслуживания, если на каждом посту за смену можно отремонтировать две ЭСУД в составе системы – «Автомобиль», а потребное количество в ремонте систем – «Автомобиль» с ЭСУД за смену – три.

Решим эту задачу, используя некоторые аспекты теории массового обслуживания и вышеописанные приемы и правила расчетов.

Итак, для нормальной работы необходимо, чтобы пропускная способность зоны текущего ремонта была больше трех систем – «Автомобиль» с ЭСУД, то есть минимальное число постов текущего ремонта должно быть два.

При оптимальном числе постов в зоне текущего ремонта суммарные убытки от простоя постов и **системы** в ожидании ремонта или технического обслуживания минимальные.

Предварительно определим вероятность поступления в ремонт четырех и более ЭСУД в составе **систем** – «Автомобиль». В соответствии с законом распределения Пуассона при среднем числе **систем** в составе автомобиля, поступающих в текущий ремонт $\lambda \Delta l$, вероятность текущего ремонта у k автомобилей

$$\text{с ЭСУД } p_k(\Delta l) = \frac{(\lambda \Delta l)^k}{k!} \cdot e^{-\lambda \Delta l}.$$

$$\text{В данном случае } \lambda \Delta l = 3, e^{-3} = 0,04979, p_k = \frac{3^k}{k!} 0,05.$$

Значения p_k в зависимости от k при среднем значении $\lambda \Delta l = 3$ ЭСУД такие:

$k \dots$	0	1	2	3	4	5
$p^k \dots$	0,05	0,15	0,225	0,225	0,169	0,101
$k \dots$	6	7	8	9	10	
$p_k \dots$	0,0506	0,0217	0,0081	0,0027	0,00081	

Вероятность, что в текущий ремонт или на техническое обслуживание поступит четыре ЭСУД в составе системы – «Автомобиль», очень большая – 0,169; с вероятностью 0,101 может быть пять систем в текущем ремонте. Вероятность, что в текущий ремонт поступит пять и больше ЭСУД в составе систем – «Автомобиль»: $p_k = 1 - \sum_{k=0}^4 p_k = 0,181$, довольно высокая.

Поэтому вероятность простоя системы – «Автомобиль» с ЭСУД в ожидании ремонта или технического обслуживания при пропускной способности двух постов четыре системы – «Автомобиль» с ЭСУД в смену весьма большая. Расчет среднего числа незанятых каналов обслуживания в зависимости от числа каналов n с пропускной способностью $\mu = 2$ каждого канала в смену при среднем поступлении ЭСУД в составе систем – «Автомобиль» в текущий ремонт или на техническое обслуживание приведен в табл. 6.6.

Таблица 6.6.

Среднее число каналов обслуживания

n	$\sum \left(n_i - \frac{k_j}{\mu} \right) p_k$	m_n
2	$0,05 \cdot 2 + 0,15 \cdot 1,5 + 0,225 \cdot 1 + 0,225 \cdot 0,5$	0,66
3	$0,05 \cdot 3 + 0,15 \cdot 2,5 + 0,225 \cdot 2 + 0,225 \cdot 1,5 + 0,17 \cdot 1 + 0,1 \cdot 0,5$	1,532
4	$0,05 \cdot 4 + 0,15 \cdot 3,5 + 0,225 \cdot 3 + 0,225 \cdot 2,5 + 0,17 \cdot 2 + 0,1 \cdot 1,5 + 0,05 \cdot 1 + 0,022 \cdot 0,5$	2,514

Расчеты вероятности простоя постов p_0 и числа ЭСУД в составе систем – «Автомобиль» m в ожидании ремонта сведены в табл. 6.7. и 6.8.

Убытки от простоя зоны текущего ремонта зависят от амортизации здания, оборудования, заработной платы рабочих и накладных расходов.

Таблица 6.7.

Вероятность простоя p_0 зоны текущего ремонта в зависимости от числа постов n при $\rho = \frac{3}{2}$

n	$\frac{\rho^k}{k!}$	$\sum_{k=0}^n \frac{\rho^k}{k!}$	$\frac{\rho^{n+1}}{n!(n-\rho)}$	$\sum_{k=0}^n \frac{\rho^k}{k!} + \frac{\rho^{n+1}}{n!(n-\rho)}$	p_0
2	$\frac{9}{8}$	$\frac{29}{8}$	$\frac{27}{8}$	$\frac{56}{8}$	0.143
3	$\frac{27}{48}$	$\frac{67}{16}$	$\frac{81}{144}$	$\frac{127,5}{24}$	0.188
4	$\frac{27}{128}$	$\frac{563}{128}$	$\frac{81}{640}$	$\frac{2896}{640}$	0.221

Таблица 6.8.

Среднее число автомобилей с ЭСУД m в ожидании ремонта в зависимости от числа постов n при $\rho = \frac{3}{2}$.

n	ρ^{n+1}	$\left(1 - \frac{\rho}{n}\right)^2$	$n \times n!$	$n \times n! \left(1 - \frac{\rho}{n}\right)^2$	$\frac{\rho^{n+1}}{n \times n! \left(1 - \frac{\rho}{n}\right)^2}$	m
2	$\frac{27}{8}$	$\frac{1}{16}$	4	$\frac{1}{4}$	13,5	1,93
3	$\frac{81}{16}$	$\frac{1}{4}$	18	$\frac{18}{4}$	$\frac{9}{8}$	0,2115
4	$\frac{243}{32}$	$\frac{25}{64}$	96	$\frac{75}{2}$	$\frac{81}{400}$	0,048

При убытке α_2 из-за простоя системы – «Автомобиль» с ЭСУД и убытки α_1 от простоя рабочего поста суммарные убытки зависят от вероятности простоя p_0 постов и среднего числа ЭСУД в составе систем – «Автомобиль» в ожидании ремонта или технического обслуживания m (табл. 6.9.).

Таблица 6.9.

Суммарные убытки от простоя зоны текущего ремонта или технического обслуживания системы – «Автомобиль» с ЭСУД в очереди в зависимости от числа постов n при $\rho = \frac{3}{2}$ при $a_1 = 1 \text{ руб/ч}$, $a_2 = 3 \text{ руб/ч}$, $\rho = \frac{3}{2}$

n	m_n	$m_n a_1$	M	$m_n a_2$	$m_n a_1 + m_n a_2$
2	0,66	0,66	1,93	5,79	6,45
3	1,532	1,532	0,2115	0,635	2,167
4	2,514	2,514	0,048	0,134	2,648

Результаты расчета табл. 6.7. и 6.8 приведены на рис. 6.9:

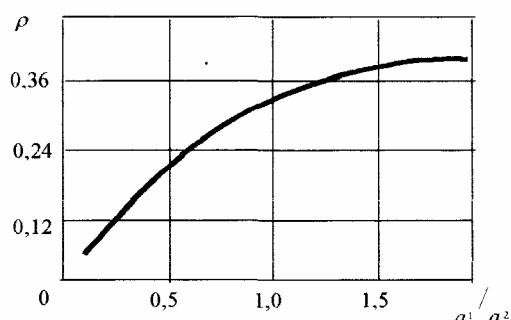


Рис. 6.9. Зависимость оптимальной приведенной плотности ρ от соотношения убытков a_1 и a_2 соответственно из-за простоя поста и системы – «Автомобиль» с ЭСУД

Как видно из табл. 6.9 и рис. 6.10, оптимальное число постов зоны текущего ремонта или технического обслуживания равно трем, при этом минимальны суммарные убытки от простоя в ожидании систем - «Автомобиль» с ЭСУД и ЭСУД в составе систем - «Автомобиль» в ожидании ремонта или технического обслуживания. Если же принять число постов два, то убытки за 1 час работы больше по сравнению с оптимальным числом постов на $6,45 - 2,167 = 4,238$ руб./ч, а за год - $2096 \times 4,238 = 9000$ руб.

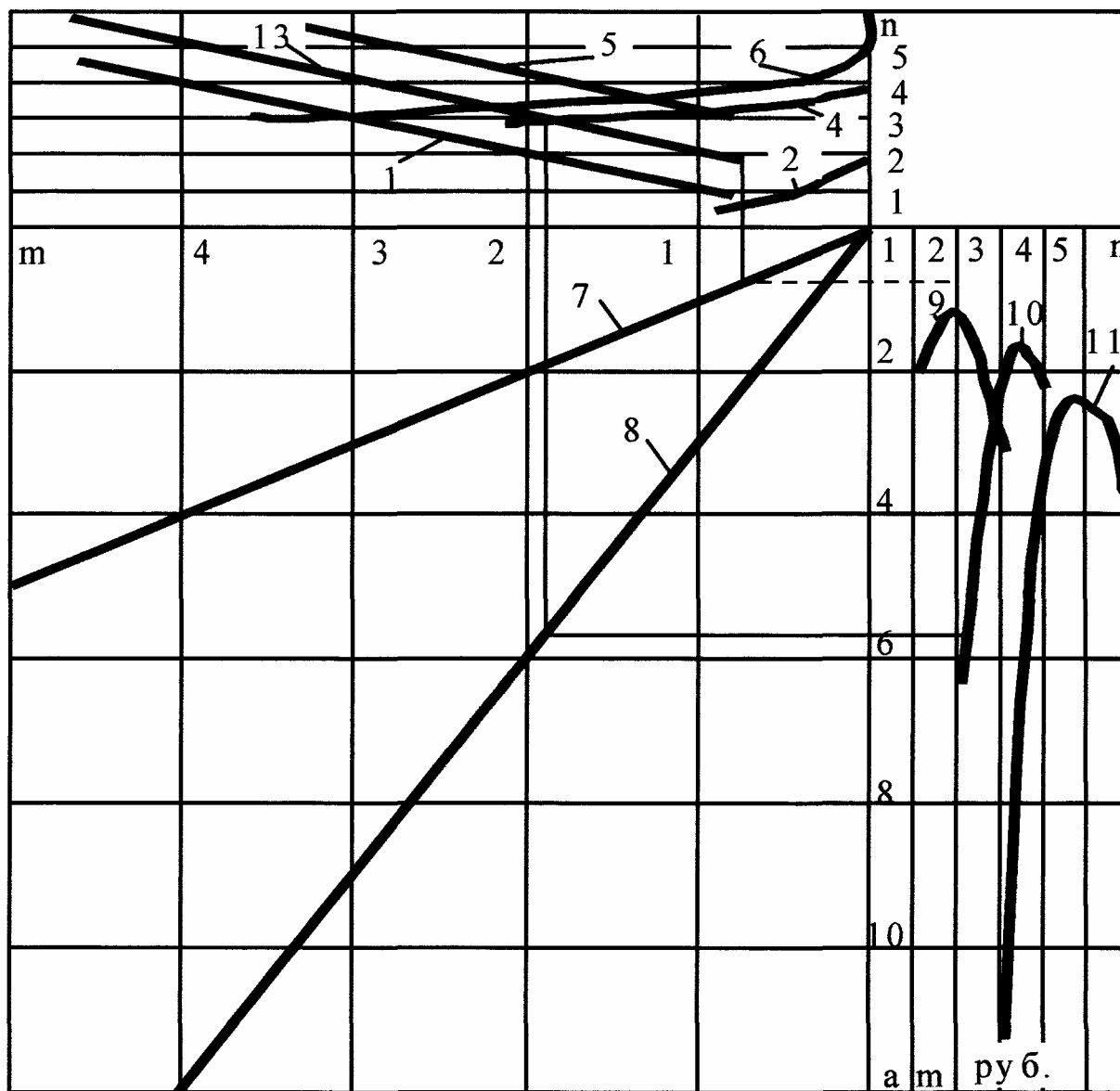


Рис. 6.10. Зависимость оптимального числа постов n от приведенной плотности p и соотношение убытков a_1 и a_2 соответственно из-за простоя постов и систем - «Автомобиль» с ЭСУД в ожидании ремонта или технического обслуживания:

1, 3, 5 - простой постов t при p соответственно $1/2$, $3/2$, $5/2$; 2, 4, 6 - среднее число систем - «Автомобиль» с ЭСУД m в ожидании ремонта технического обслуживания при p соответственно $1/2$, $3/2$, $5/2$; 7 - при $a_1 = 1$ руб; 8 - $a_2 = 3$ руб; 9, 10, 11 - суммарные убытки от простоя постов и систем - «Автомобиль» с ЭСУД при p соответственно $1/2$, $3/2$, $5/2$

При помощи зависимости (рис.6.10) оптимального числа постов от приведенной плотности и соотношения убытков из-за простоя постов и систем – «Автомобиль» с ЭСУД в ожидании ремонта можно определить оптимальное число постов при заданных приведенной плотности и убытках из-за простоя постов и систем – «Автомобиль» с ЭСУД.

Покажем это на примере, когда $Q = \frac{3}{2}$, $a_1 = 1 \text{ руб.}$, $a_2 = 3 \text{ руб.}$. Вначале выясним, сколько в среднем простаивает постов при $\rho = \frac{3}{2}$. Прямая 3 характеризует простой постов m_n при $\rho = \frac{3}{2}$; при $n = 2$ простой поста меньше единицы, то есть меньше одного поста, убытки при этом (при стоимости простоя поста 1 руб.) составляют меньше 1 руб. Об этом можно судить по пересечению перпендикуляра с ординатой am из точки пересечения прямой 7 (убытки 1 руб.) с перпендикуляром из абсциссы m при $n = 2$ и $\rho = \frac{3}{2}$; значение am переносят до ординаты am при $n = 2$ (правый нижний сектор). Убытки от простоя систем – «Автомобиль» с ЭСУД при числе постов $n = 2$ определяют аналогично, только вместо прямой 3 принимают кривую 4 (среднее число ЭСУД на обслуживание в составе систем – «Автомобиль» в очереди при $\rho = \frac{3}{2}$ и $n = 2$) и перпендикуляр к оси m опускают до пересечения с прямой 8 (удельные убытки 3 руб.), а затем из этой точки опускают перпендикуляр к оси am до пересечения с параллельной осью при $n = 2$. На этой параллельной оси am откладывают всю сумму убытков от простоя и постов и систем – «Автомобиль» с ЭСУД (на оси am при $n = 2$ отмечено точкой). Аналогично определяют $\sum am$ при $n = 3 \dots 4$ и $\rho = \frac{3}{2}$; при других $\rho (\frac{1}{2}, \frac{5}{2})$ используют прямую 1 и кривую 2, прямую 5 и кривую 6 соответственно при $n = 1, \dots, 6$.

Приведенная методика определения оптимального числа постов текущего ремонта справедлива для разомкнутой системы массового обслуживания. При этом следует учитывать не только случайность периодичности и трудоемкости ремонта, но и закономерность изменения удельной трудоемкости текущего ремонта по мере увеличения пробега системы – «Автомобиль» с ЭСУД с начала эксплуатации.

Таким образом, оптимальным планированием ремонта создают условия, при которых минимальные суммарные удельные затраты или максимальные доходы или максимальная производительность при работе ЭСУД в составе системы – «Автомобиль» в заданных эксплуатационных условиях, являются **пассивной формой оптимизации**.

Методики оптимизации режима технического обслуживания системы – «Автомобиль» с ЭСУД и организации перевозок при минимальной себестоимости в принципе одинаковы и поэтому можно ограничиться **методикой опти-**

мизации условий поддержания ЭСУД в составе системы - «Автомобиль» в работоспособном состоянии.

Управлять процессом, направлять его в нужном русле - идеальный случай. В общем виде качество управления можно оценить затратами в заданных, неподвластных условиях. Для этого нужно иметь хорошую информацию о развитии процесса в прошлом, закономерность развития, умение предвидеть. Закономерность связывает подсистемы в единую систему при помощи математических зависимостей и позволяет определить количественные значения параметров эффективности работы электронных систем управления двигателями автомобилей.

Таким образом, для управления необходимо прежде всего знать математическую модель процесса и параметры этой модели, значения которых определяют на основе информации о развитии процесса за прошедшие эволюционные периоды и время.

ЗАКЛЮЧЕНИЕ

Таким образом, выше рассмотрены ряд методов анализа **оптимального управления** в процессах и системах проектирования и управления, что позволяет сделать вывод и дает основание полагать, что все в «природе» может быть описано определенными законами **математики**, что все процессы развития управления, **процессов и систем управления** моделируются математическими и электронными моделями. Все процессы и **управление**, равно как автоматическое и автоматизированное, системы и процессы как технологические и экономические (так и социальные) подчиняются принципу иерархичности построения с упорядоченной взаимосвязкой и увязкой всех компонентов и элементов определенными правилами и требованиями к ним для более эффективного функционирования с позиции функционально- стоимостного анализа, и критериев **функциональности и стоимости**. То есть если применить **проверки сравнением** (верификацию) процессов и систем управления, с использованием основных методов и законов классической высшей математики и основных критериев **функциональности**, то функциональность можно рассматривать как единство требований, то есть **действенности и целесообразности; качества и надежности; полезности** и эффективности, а также **оптимизации затрат** (или стоимости) и, наконец, получение оптимального продукта проектирования и управления разработками и т.д. Задача критического сопоставления верификационных методов анализа оптимального управления и принятия оптимальных технических решений в настоящей работе не стояла, так как здесь необходимо было выяснить и сравнить уже имеющиеся методы анализа, использовать их при принятии проектных решений с соответствующей интерпретацией, и проверить на некоторых примерах, например, на электронных системах управления двигателями (ЭСУД) их жизнеспособность и эффективность, поскольку в настоящее время **оптимизация** именно **проверок сравнением**, то есть выбора оптимального метода анализа - **велика**.

Многочисленные исследования и практическое применение верификационных методов анализа оптимального проектирования и управления разработками, например, авиационной и другой техники, показали, что нет универсальных методов (абсолютно универсальных) и быть не может, что каждый метод **анализа** должен соответствовать решаемой задаче в своей предметной области. Поэтому здесь авторы обобщили огромный опыт отечественных и зарубежных ученых по использованию **верификации** методов анализа, и дали свой комплексный подход к решению задач по определению оптимального управления и оптимизации

принимаемых решений в процессах и системах управления в любой предметной области производства, технического сервиса и обслуживания техники.

Работа будет полезна студентам технических вузов, изучающим организацию процессов и систем проектирования и управления, а также систем автоматизации и автоматизированных систем управления и т.д.; инженерным работникам в сфере проектирования, сервиса и обслуживания техники; аспирантам, преподавателям технических вузов, колледжей и лицеев в области авиаракето- и автомобилестроения, и др.

Современная система образования, особенно в технических вузах, предусматривает ориентацию будущих специалистов в большей степени на организацию технических систем гражданской ориентации, хотя и не снижает акцента по изучению специальных систем, ориентированных на производство военной техники.

В этой связи, знакомство с настоящей работой, дает возможность специалистам рассматривать широкий круг проблем по оптимизации управления, а тем более изучению методов проверок сравнением (верификации) различных методов оптимизации широкого круга технических (автоматических и автоматизированных) систем на основе классической высшей математики и функционально-стоимостной инженерии по критериям «функциональности», функции (как полезного действия, свойства или состояния) и стоимости, комплексных показателей эффективности технических и управленческих систем.

Авторы выражают искреннюю благодарность рецензентам и научному редактору за ценные замечания, которые в процессе вёрстки настоящей работы были учтены и устранены.

СПИСОК ЛИТЕРАТУРЫ

1. Гельфанд И.М., Фомин С.В. Вариационное исчисление. М.: Физматгиз, 1961.
2. Влчек Р. Функционально-стоимостной анализ в управлении: Сокр. пер. с чеш. - М.: Экономика, 1986. - 176 с.
3. Понtryгин Л.С., Болтянский В.Г. и др. Математическая теория оптимальных процессов. М.: Физматгиз, 1961.
4. Справочное пособие по теории систем автоматического регулирования и управления / Под общ. ред. Е.А. Санковского. - Мн.: «Вышэйш. школа», 1973.-584 с.
5. Автоматизация мелкосерийного машиностроительного производства и качества продукции / Р.И. Адгамов, В.М. Белоноч, Ю.Н. Блощицын и др.; под ред. Р.И. Адгамов. - М.: Машиностроение, 1983. - 280 с.
6. Егор С.М. и др. Основы автоматизированного проектирования самолетов: Уч. пособие для студентов авиационных специальностей вузов. / С.М. Егор, Н.К. Лисейцев, О.С. Самойлович. - М.: Машиностроение, 1986.-232 с.
7. Аворгенков В.И. и др. САПР технологических процессов, приспособлений и режущих инструментов: Учебное пособие для вузов / В.И. Авер-генков, И.А. Каштальян, А.П. Пархутик. - Мн.: «Вышэйш. школа», 1993. -288с.
8. Быков В.П. Методическое обеспечение САПР в машиностроении. - Л.: Машиностроение, Ленингр. отд - ние, 1989. - 255 с.
9. Проектирование систем автоматизации технологических процессов: Справочное пособие / А.С. Ключев; ред. А.С. Ключев. - 2-е изд., перераб. и доп. - М.: Энергоиздат., 1990. - 464 с.
10. Основы теории автоматизированного управления: Учебник для авиационных вузов / В.С. Булыгин, Ю.С. Гришанин, Н.Б. Судзиловский и др.; под ред. Н.Б. Судзиловского. - М.: Машиностроение, 1985. - 512 с.
11. Бержеев М.М., Заляев И.А. и др. Основы системы автоматизированного проектирования: Учебное пособие: - Издательство Казанского университета, 1988.- 254 с.
12. Автоматизация процессов подготовки авиационного производства на базе ЭВМ и оборудования с ЧПУ / В. А. Вайсбург, Б.А. Медведев, А.Н.Бакушский и др. - М.: Машиностроение, 1985. -216с.
13. Вальнов В.М., Верминин В.Е. Автоматизированные системы управления технологическими процессами. - 3-е изд., перераб. и доп. - Л.: Политехнического университета, 1991. - 269 с.
14. Энкарначчо Ж., Шлехтендаль Э. Автоматизированное проектирование. Основные понятия и архитектура систем: Пер. с англ. - М.: Радио и связь, 1986.-288 с.

15. Татарникова Е.Н. Нормативно-справочное обеспечение АСУ машиностроительного предприятия. - Л.: Машиностроение, Ленингр. отд- ния, 1978.-160 с.
16. Бойков А. Д. Расчет систем автоматического управления с использованием вычислительных машин.: Учебное пособие: - Морд. госуд. унив. им. И.П.Огарева, 1980.-174 с.
17. Шибалов Т.П. и др. Автоматизация испытаний и контроля авиационных ГТД. - М.: Машиностроение, 1987. - 280 с.
18. Смехов А.А. Автоматизированные склады .- 3-е изд., перераб. и доп. -М.: Машиностроение, 1979. - 288 с.
19. Куропаткин П.В. Теория автоматического управления.: Учебное пособие для электротехн. специальностей вузов. - М.:«Высшая школа», 1983. -528с.
20. Роботизированные производственные комплексы, / Ю.Г. Козорев, А.А. Куринов и др. - М: Машиностроение, 1987. - 272 с. (Автоматические манипуляторы и робототехнические системы).
21. Барангукова И.М., Гусев А.А. и др. Проектирование технологии: Учебник для студентов машиностроительных специальностей вузов / И.М. Барангукова, А.А.Гусев и др. ; под общ. ред. Ю.М. Соломенцева. -М.: Машиностроение, 1990. - 416 с. (Технология автоматизированного машиностроения).
22. Дудорин В. И. Моделирование в задачах управления производством. -М.: Статистика, 1990. - 232 с.
23. Парамонов Ф.И. Моделирование процессов производства. - М.: Машиностроение, 1994. -232 с.
24. Вендров А.М. CASE — технологически современные методы и средства проектирования информационных систем. - М.: Финансы и статистика, 1998. - 176 с.
25. Колганов И.М. Сборочные работы при производстве широкофюзеляжных самолетов. Технологические процессы, выбор варианта: Учебное пособие. - : Ульяновск, УлГТУ, 1999. - 96 с.
26. Экономика и математические методы, том 34, выпуск 4, октябрь - декабрь 1998, РАН ЦЭМИ, ИПР.
27. Попов Е.С., Брыков Н.Н., Пугачёв Г.А. К методике исследования сопротивляемости материалов изнашивания // Заводская лаборатория. - 1984. -№ 11.-С.76-78.
28. Розоноэр Л.И. Принцип максимума Л.С. Понтрягина в теории относительных систем I, II, III. - Автоматика и техника, 1959, №№10,11, 12.
29. Ляпунов А.М. Общая задача об устойчивости движения. М.: - Л.: Гос. изд-во техн.-теорией лит. 1950
30. Курант Р., Гильберт Д. Методы математической физики, Т.I. М. - Л.: Гостеиздат, 1951.

31. Канторович Л.В., Крылов В.И. Приближенные Методы высшего анализа. М.: Физматгиз, 1962.
32. Гантмахер Ф.Р. Теория матриц. Изд-е 3-е. М.: наука, 1967.
33. Колмогоров А.Н., Фомин С.В. Элементы теории функций и функционального анализа. Изд-е 2-е. М.: Наука, 1968.
34. Колмогоров А.Н., Основные понятия теории вероятностей. Изд-е второе. М.: Наука, 1974.
35. Сейж Э.П., Уайт Ч.С., Ш Оптимальное управление системами: Пер. с англ. / Под ред. Б.Р. Левана. - М.: Радио и связь, 1982. - (Второе изд.: США, 1977).
36. Агеев В.М. Павлова Н.В. Сергейчик В.А. Автоматизация проектирования бортовых информационно-измерительных систем: Учеб. пособие / под ред. В.В. Петрова. М.: изд. МАИ, 1985.
37. Голинкевич Т.А. прикладная теория надежности: Учебник для вузов. М.: **Высшая школа**, 1977.
38. Костюк В.И., Ходаков В.Е. Системы отображения информации и инженерная психология: Учеб. пособие для вузов. Киев: Высшая школа, 1977.
39. Садыков Ф.Р. Автоматизированное проектирование систем управления летных аппаратов. М.: изд. МАИ, 1985.
40. Балакшин Б.С. Основы технологии машиностроения, М.: Машиностроение, 1969.
41. Павлов В.В. Математическое обеспечение САПР в производстве летательных аппаратов М.: изд. МФТИ, 1978.
42. Корсоков В.С. Автоматизация производственных процессов. М.: Высшая школа, 1978.
43. Павлов В.В. Теоретические основы сборки Л А. М.: изд. МФТИ, 1981.
44. Автоматизированные системы технологической подготовки производства в машиностроении. / Под ред. Г.К. Горанского. М.: Машиностроение, 1976.
45. Попов П.М., Ляшко Ф.Е. Оптимальное управление в ходе эволюционного развития процессов и систем: Учеб. пособие. - Ульяновск, 2000.
46. Катыс Г.П. Информационные сканирующие системы. М.: «Машиностроение», 1965.
47. Попов П.М. Некоторые правила и приемы определения устойчивости линейных звеньев в отдельных системах автоматического управления Л А. Учеб. пособие. - Ульяновск, 2000.
48. Попов П.М. Оптимизация технических решений проектирования и управления на основе экономико-математических методов анализа. -Ульяновск: УлГТУ, 2000.
49. Бабушкин А.И. Моделирование и оптимизация сборки Л А. - М.: Машиностроение, 1990.
50. Авдоськин Ф.Н. Оптимизация изменения технического состояния автомобиля в процессе эксплуатации. - М.: Транспорт, 1993.

51. Аринин И.Н. Диагностирование технического состояния автомобилей. М.: Транспорт, 1978.
52. Бухарин Н.А. Котиков Ю.Г. и др. О приведении нестационарного случайного процесса к стационарному при исследовании динамических нагрузок//Двигателестроение. 1975.
53. Исаева Л.С. Оптимальные сроки службы автомобилей. М.: Транспорт, 1976.
54. Ланге О. Оптимальные решения. М.: Прогресс, 1967.
55. Луйк И.А. Теоретические основы планирования технической эксплуатации машинного парка. Киев. Вища школа, 1976.
56. Попов Е.С. Брыков Н.Н. и др. К методике исследования сопротивляемости материалов изнашиванию // Заводская лаборатория. 1984.
57. Сергеев А.Г. Точность и достоверность диагностики автомобилей. М.: транспорт, 1980.
58. Хрущев М.М. Бабичев М.А. Исследование изнашивания металлов. М.: Изд-во АН СССР, 1960.
59. Шушик С.В. Основы технической эксплуатации автомобилей. Минск: Вышэйша школа, 1981.
60. Ярковец А.И. Основы механизации и автоматизации технологических процессов в самолетостроении. М.: Машиностроение, 1981.
61. Кибернетика и системный анализ. Международный научно-технический журнал. Инст. Киберн. им. В.М. Глушкова, НАН Украины, 2000.
62. Экономика и математические методы. - М.: изд-во «Наука», 2000.

Научное издание

ПОПОВ Петр Михайлович

ПОПОВ Сергей Петрович

**ВЕРИФИКАЦИОННЫЕ МЕТОДЫ АНАЛИЗА ОПТИМАЛЬНОГО
УПРАВЛЕНИЯ ПРОЦЕССАМИ И СИСТЕМАМИ**

Изд. лиц. 020640 от 22.10.97. Подписано в печать 20.04.01. Формат 60х84/16. Бумага писчая. Печать офсетная. Усл.п.л. 11,28. Уч.-изд.л. 10,80, Тираж 110 экз. Заказ 1509

Ульяновский государственный технический университет

432027, г. Ульяновск, Сев. Венец, 32.

Типография УлГТУ. 432027, Ульяновск, ул. Сев. Венец, 32.